

Efficient Estimation of Partially Linear Models for Spatial Data over Complex Domains

Li Wang^a, Guannan Wang^b, Min-Jun Lai^c and Lei Gao^a

^aIowa State University, ^bCollege of William & Mary and ^cUniversity of Georgia

Abstract: In this paper, we study the estimation of partially linear models for spatial data distributed over complex domains. We use bivariate splines over triangulations to represent the nonparametric component on an irregular two-dimensional domain. The proposed method is formulated as a constrained minimization problem which does not require constructing finite elements or locally supported basis functions. Thus, it allows an easier implementation of piecewise polynomial representations of various degrees and various smoothness over an arbitrary triangulation. Moreover, the constrained minimization problem is converted into an unconstrained minimization via a QR decomposition of the smoothness constraints, which leads to a penalized least squares method to estimate the model. The estimators of the parameters are proved to be asymptotically normal under some regularity conditions. The estimator of the bivariate function is consistent, and its rate of convergence is also established. The proposed method enables us to construct confidence intervals and permits inference for the parameters. The performance of the estimators is evaluated by two simulation examples and by a real data analysis.

Key words and phrases: Bivariate splines, Penalty, Semiparametric regression, Spatial data, Triangulation.

1. Introduction

In many geospatial studies, spatially distributed covariate information is available. For example, geographic information systems may contain measurements obtained from satellite images at some locations. These spatially explicit data can be useful in the construction and estimation of regression models, but sometimes they are distributed over irregular two-dimensional (2-D) domains that may have complex boundaries or holes inside. It is well known that many conventional smoothing tools suffer from the problem of “leakage” across the complex domains, which refers to the poor estimation over difficult regions by smoothing inappropriately across boundary features, such as peninsulas; see excellent discussions in Ramsay (2002) and Wood, Bravington and Hedley (2008). In this paper, we propose to use bivariate splines (smooth piecewise polynomial functions over a triangulation of the domain of interest) to model spatially explicit datasets which enable us to overcome the “leakage” problem and find the estimation more accurately.

Address for correspondence: Li Wang, Department of Statistics and the Statistical Laboratory, Iowa State University, Ames, IA, USA. Email: lilywang@iastate.edu

We focus here on the partially linear model (PLM), popularized by Hardle, Liang and Gao (2000), for data randomly distributed over 2-D domains. To be more specific, let $\mathbf{X}_i = (X_{i1}, X_{i2})^T$ be the location of i -th point, $i = 1, \dots, n$, which ranges over a bounded domain $\Omega \subseteq \mathbb{R}^2$ of arbitrary shape, for example, a domain with polygon boundary. Let Y_i be the response variable and $\mathbf{Z}_i = (Z_{i1}, \dots, Z_{ip})^T$ be the predictors at location \mathbf{X}_i . Suppose that $\{(\mathbf{Z}_i, \mathbf{X}_i, Y_i)\}_{i=1}^n$ satisfies the following model

$$Y_i = \mathbf{Z}_i^T \boldsymbol{\beta} + g(\mathbf{X}_i) + \epsilon_i, \quad i = 1, \dots, n, \quad (1.1)$$

where $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^T$ are unknown parameters, $g(\cdot)$ is some unknown but smooth bivariate function, and ϵ_i 's are i.i.d random noises with $E(\epsilon_i) = 0$ and $\text{Var}(\epsilon_i) = \sigma^2$. Each ϵ_i is independent of \mathbf{X}_i and \mathbf{Z}_i . In many situations, our main interest is in estimating and making inference for the regression parameters $\boldsymbol{\beta}$, which provides measures of the effect of the covariate \mathbf{Z} after adjusting for the location effect of \mathbf{X} .

If $g(\cdot)$ is a univariate function, model (1.1) becomes a typical PLM. In the past three decades, flexible and parsimonious PLMs have been extensively studied and widely used in many statistical applications, from biostatistics to econometrics, from engineering to social science; see Chen, Liang and Wang (2011), Huang, Zhang and Zhou (2007), Liang and Li (2009), Liu, Wang and Liang (2011), Ma, Song and Wang (2013), Ma and Yang (2011), Wang, et al (2011), Wang, et al (2014), Zhang, Cheng and Liu (2011) for some recent works on PLMs. When $g(\cdot)$ is a bivariate function, there are two popular estimation tools: bivariate P-splines Marx and Eilers (2005) and thin plate splines Wood (2003). Later, Xiao, Li and Ruppert (2013) proposed an efficient sandwich smoother, which has a tensor product structure that simplifies an asymptotic analysis and can be fast computed. Their method has been applied to quantifying the lifetime circadian rhythm of physical activity Xiao, et al (2015). The application to spatial data analysis over complex domains, however, has been hampered due to the scarcity of bivariate smoothing tools that are not only computationally efficient but also theoretically reliable to solve the problem of “leakage” across the domain. Traditional smoothing methods in practical data analysis, such as kernel smoothing, wavelet-based smoothing, tensor product splines and thin plate splines, usually perform poorly for those data, since they do not take into account the shape of the domain and also smooth across concave boundary regions.

There are several challenges when going from rectangular domains to irregular domains with complex boundaries or holes. Some efforts have recently been devoted to studying the smoothing over irregular domains, and significant progress has been made. Most of them are based on the roughness penalization approach Green and Silverman (1994). To deal with irregular domains, Wang and Ranalli (2007) applied low-rank thin-plate splines defined as functions of the geodesic distance instead of the Euclidean distance. Eilers (2006) utilized the

Schwarz-Christoffel transform to convert the complex domains to regular domains. To solve the “leakage” problem, in a pioneering paper, Ramsay (2002) suggested a penalized least squares approach with a Laplacian penalty and transformed the problem to that of solving a system of partial differential equations (PDEs). Wood, Bravington and Hedley (2008) provided an elegant solution and developed the soap film smoothing estimator for smoothing over difficult regions that can be represented by a low-rank basis and one or two quadratic penalties. Recently, Sangalli, Ramsay and Ramsay (2013) extended the method in Ramsay (2002) to the PLMs, which allows for spatially distributed covariate information to be in the models. The data smoothing problem in Sangalli, Ramsay and Ramsay (2013) is solved using finite element method (FEM), a method mainly developed and used to solve PDEs. Although their method is practically useful, the theoretical properties of the estimation are not investigated.

In this paper, we tackle the estimation problem differently from Sangalli, Ramsay and Ramsay (2013). Instead of using FEM, we approximate the nonparametric function $g(\cdot)$ by using the bivariate splines over triangulations in Lai and Schumaker (2007). An important feature of this approach is that it uses splines for applications without constructing locally supported splines or finite elements and without computing the dimension. This method has been shown to be more efficient and flexible than the conventional FEM in data fitting problems and solving PDEs; see Awanou, Lai and Wenston (2005), Ettinger, Guillas and Lai (2015), Guillas and Lai (2010), Lai and Wang (2013) and Liu, Guillas and Lai (2015). For example, the users can choose spline functions of flexible degrees and various smoothness across any given domain. Another advantage is that the linear systems arising in this approach are more easy to assemble than those from the finite elements or locally supported spline basis functions. The linear systems are sparser than that from any macro-FEM method. In addition, due to the great scalability, the assembling can be computed in parallel.

To the best of the authors’ knowledge, statistical aspects of smoothing for PLMs by using bivariate splines have not been discussed in the literature so far. This paper presents the first attempt at investigating the asymptotic properties of the PLMs for data distributed on a non-rectangular complex region. We study the asymptotic properties of the least squares estimators of β and $g(\cdot)$ by using bivariate splines over triangulations with a penalty term. We show that our estimator of β is root- n consistent and asymptotically normal, although the convergence rate of the estimator of the nonparametric component $g(\cdot)$ is slower than root- n . A standard error formula for the estimated coefficients is provided and tested to be accurate enough for practical purposes. Hence, the proposed method enables us to construct confidence intervals for the regression parameters. We also obtain the convergence rate for the functional estimator of $g(\cdot)$. We show, by using numerical studies, that our method is competitive with existing methods such as the soap lm smoother and the thin-plate regression spline.

The rest of the paper is organized as follows. In Section 2, we give a brief review of the triangulations and propose our estimation method based on penalized bivariate splines. We also discuss the details on how to choose the penalty parameters. Section 3 is devoted to the asymptotic analysis of the proposed estimators. Section 4 provides a detailed simulation to compare several methods in two different scenarios and explores the estimation and prediction accuracy. Section 5 studies a real dataset on house values over California. Some concluding remarks are given in Section 6. Technical details are provided in Appendix A.

2. Triangulations and Penalized Spline Estimators

Our estimation method is based on penalized bivariate splines over triangulations. The idea is to approximate the function $g(\cdot)$ by bivariate splines that are piecewise polynomial functions over a 2D triangulated domain. We use this approximation to construct least squares estimators of the linear and nonlinear components of the model with a penalization term. In the following of this section, we describe the background of triangulations, B-form bivariate splines and introduce the penalized spline estimators.

2.1. Triangulations

Triangulation is an effective strategy to handle data distribution on irregular regions with complex boundaries and/or interior holes. Recently, it has attracted substantial recent attention in many applied areas, such as geospatial studies, numerical solutions of PDEs, image enhancements, and computer aided geometric design. See, for example, the recent comprehensive book by Lai and Schumaker (2007) and the article by Lai (2008).

We use τ to denote a triangle which is a convex hull of three points not located in one line. A collection $\Delta = \{\tau_1, \dots, \tau_N\}$ of N triangles is called a triangulation of $\Omega = \cup_{i=1}^N \tau_i$ provided that if a pair of triangles in Δ intersect, then their intersection is either a common vertex or a common edge. In general, any kind of polygon shapes can be used for the partition of Ω . In this paper we consider triangulations of Ω because any polygonal domain of arbitrary shape can be partitioned into finitely many triangles. In the following, we assume that all \mathbf{X}_i s are inside triangles of Δ . That is, they are not on edges or vertices of triangles in Δ . Otherwise, we can simply count them twice or multiple times if any observation is located on an edge or at a vertex of Δ . Given a triangle $\tau \in \Delta$, we let $|\tau|$ be its longest edge length, and denote the size of Δ by $|\Delta| = \max\{|\tau|, \tau \in \Delta\}$, i.e., the length of the longest edge of Δ . Furthermore, let ρ_τ be the radius of the largest circle inscribed in τ . We measure the quality of a triangulation Δ by $\delta_\Delta = \max_{\tau \in \Delta} |\tau|/\rho_\tau < \infty$, which is equivalent to the smallest angle of Δ . The study in Lai and Schumaker (2007) shows that the approximation of spline spaces over Δ is dependent on δ_Δ , i.e., the larger the δ_Δ is, the worse the spline approximation is. In the rest of the paper, we restrict our attention to the triangulations satisfying $\delta_\Delta < \delta$ for a positive constant δ .

2.2. B-form bivariate splines

In this section we give a brief introduction to the bivariate splines. More in-depth description can be found in Lai and Schumaker (2007), Lai (2008), as well as Zhou and Pan (2014) and the details of the implementation is provided in Awanou, Lai and Wenston (2005). Let $\tau = \langle \mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3 \rangle$ be a non-degenerate (i.e. with non-zero area) triangle with vertices \mathbf{v}_1 , \mathbf{v}_2 , and \mathbf{v}_3 . Then for any point $\mathbf{v} \in \mathbb{R}$, there is a unique representation in the form

$$\mathbf{v} = b_1 \mathbf{v}_1 + b_2 \mathbf{v}_2 + b_3 \mathbf{v}_3$$

with $b_1 + b_2 + b_3 = 1$, where b_1 , b_2 and b_3 are called the barycentric coordinates of the point \mathbf{v} relative to the triangle τ . The Bernstein polynomials of degree d relative to triangle τ is defined as

$$B_{ijk}^{\tau,d}(\mathbf{v}) = \frac{d!}{i!j!k!} b_1^i b_2^j b_3^k.$$

Then for any $\tau \in \Delta$, we can write the polynomial piece of spline s restricted on $\tau \in \Delta$ as

$$s|_{\tau} = \sum_{i+j+k=d} \gamma_{ijk}^{\tau} B_{ijk}^{\tau,d},$$

where the coefficients $\gamma_{\tau} = \{\gamma_{ijk}^{\tau}, i+j+k=d\}$ are called B-coefficients of s .

For a nonnegative integer r , let $\mathbb{C}^r(\Omega)$ be the collection of all r -th continuously differentiable functions over Ω . Given a triangulation Δ , let $\mathbb{S}_d^r(\Delta) = \{s \in \mathbb{C}^r(\Omega) : s|_{\tau} \in \mathbb{P}_d(\tau), \tau \in \Delta\}$ be a spline space of degree d and smoothness r over triangulation Δ , where \mathbb{P}_d is the space of all polynomials of degree less than or equal to d . For notation simplicity, let $\mathbb{S} = \mathbb{S}_{3r+2}^r(\Delta)$ for a fixed smoothness $r \geq 1$, and we know that such a spline space has the optimal approximation order (rate of convergence) for noise-free datasets; see Lai and Schumaker (1998) and Lai and Schumaker (2007).

For notation simplicity, let $\{B_{\xi}\}_{\xi \in \mathcal{K}}$ be the set of degree- d bivariate Bernstein basis polynomials for \mathbb{S} , where \mathcal{K} stands for an index set of K Bernstein basis polynomials. Then for any function $s \in \mathbb{S}$, we can represent it by using the following basis expansion:

$$s(\mathbf{x}) = \sum_{\xi \in \mathcal{K}} B_{\xi}(\mathbf{x}) \gamma_{\xi} = \mathbf{B}(\mathbf{x})^T \boldsymbol{\gamma}, \quad (2.1)$$

where $\boldsymbol{\gamma}^T = (\gamma_{\xi}, \xi \in \mathcal{K})$ is the spline coefficient vector. To meet the smoothness requirement of the splines, we need to impose some linear constraints on the spline coefficients $\boldsymbol{\gamma}$ in (2.1). We require that $\boldsymbol{\gamma}$ satisfies $\mathbf{H}\boldsymbol{\gamma} = 0$ with \mathbf{H} being the matrix for all smoothness conditions across shared edges of triangles, which depends on r and the structure of the triangulation.

2.3. Penalized Spline Estimators

To define the penalized spline method, for any direction x_j , $j = 1, 2$, let $D_{x_j}^q f(\mathbf{x})$ denote the q -th order derivative in the direction x_j at the point $\mathbf{x} = (x_1, x_2)$. Let

$$\mathcal{E}_v(f) = \sum_{\tau \in \Delta} \int_{\tau} \sum_{i+j=v} \binom{v}{i} (D_{x_1}^i D_{x_2}^j f)^2 dx_1 dx_2 \quad (2.2)$$

be the energy functional for a fixed integer $v \geq 1$. Although all partial derivatives up to the chosen order v can be included in the penalty of (2.2), for simplicity, in the remaining part of the paper, we use $v = 2$, and one can study the similar problem for general $v \geq 2$. When $v = 2$,

$$\mathcal{E}_2(f) = \int_{\Omega} ((D_{x_1}^2 f)^2 + 2(D_{x_1} D_{x_2} f)^2 + (D_{x_2}^2 f)^2) dx_1 dx_2, \quad (2.3)$$

which is similar to the thin-plate spline penalty (Green and Silverman, 1994) except the latter is integrated over the entire plane \mathbb{R}^2 . Sangalli, Ramsay and Ramsay (2013) used a different roughness penalty from (2.3), specifically, they use the integral of the square of the Laplacian of f , that is, $\lambda \int_{\Omega} (D_{x_1}^2 f + D_{x_2}^2 f)^2 dx_1 dx_2$. Both forms of penalties are invariant with respect to Euclidean transformations of spatial co-ordinates, thus, the bivariate smoothing does not depend on the choice of the coordinate system.

Given $\lambda > 0$ and $\{(\mathbf{Z}_i, \mathbf{X}_i, Y_i)\}_{i=1}^n$, we consider the following minimization problem:

$$\min_{s \in \mathbb{S}} \sum_{i=1}^n \{Y_i - \mathbf{Z}_i^T \boldsymbol{\beta} - s(\mathbf{X}_i)\}^2 + \lambda \mathcal{E}_v(s). \quad (2.4)$$

Let $\mathbf{Y} = (Y_1, \dots, Y_n)^T$ be the vector of n observations of the response variable. Denote by $\mathbf{X}_{n \times 2} = \{(X_{i1}, X_{i2})\}_{i=1}^n$ the location design matrix and $\mathbf{Z}_{n \times p} = \{(Z_{i1}, \dots, Z_{ip})\}_{i=1}^n$ the collection of all covariates. Denote by \mathbf{B} the $n \times K$ evaluation matrix of Bernstein basis polynomials whose i -th row is given by $\mathbf{B}_i^T = \{B_{\xi}(\mathbf{X}_i), \xi \in \mathcal{K}\}$. Then the minimization problem in (2.4) reduces to

$$\min_{\boldsymbol{\beta}, \boldsymbol{\gamma}} L(\boldsymbol{\beta}, \boldsymbol{\gamma}) = \min_{\boldsymbol{\beta}, \boldsymbol{\gamma}} \{\|\mathbf{Y} - \mathbf{Z}\boldsymbol{\beta} - \mathbf{B}\boldsymbol{\gamma}\|^2 + \lambda \boldsymbol{\gamma}^T \mathbf{P} \boldsymbol{\gamma}\} \quad \text{subject to} \quad \mathbf{H}\boldsymbol{\gamma} = \mathbf{0}, \quad (2.5)$$

where \mathbf{P} is the block diagonal penalty matrix satisfying that $\boldsymbol{\gamma}^T \mathbf{P} \boldsymbol{\gamma} = \mathcal{E}_v(\mathbf{B}\boldsymbol{\gamma})$.

To solve the constrained minimization problem (2.5), we first remove the constraint via QR decomposition of the transpose of the constrain matrix \mathbf{H} . Specifically, we write

$$\mathbf{H}^T = \mathbf{Q}\mathbf{R} = (\mathbf{Q}_1 \mathbf{Q}_2) \begin{pmatrix} \mathbf{R}_1 \\ \mathbf{R}_2 \end{pmatrix}, \quad (2.6)$$

where \mathbf{Q} is an orthogonal matrix and \mathbf{R} is an upper triangle matrix, the submatrix \mathbf{Q}_1 is the first r columns of \mathbf{Q} , where r is the rank of matrix \mathbf{H} , and \mathbf{R}_2 is a matrix of zeros. It is easy to see the following result.

Lemma 1. *Let $\mathbf{Q}_1, \mathbf{Q}_2$ be submatrices as in (2.6). Let $\boldsymbol{\gamma} = \mathbf{Q}_2\boldsymbol{\theta}$ for a vector $\boldsymbol{\theta}$ of appropriate size. Then $\mathbf{H}\boldsymbol{\gamma} = \mathbf{0}$. On the other hand, if $\mathbf{H}\boldsymbol{\gamma} = \mathbf{0}$, then there exists a vector $\boldsymbol{\theta}$ such that $\boldsymbol{\gamma} = \mathbf{Q}_2\boldsymbol{\theta}$.*

Proof. By (2.6), we have $\mathbf{H}^\mathbf{T} = \mathbf{Q}_1\mathbf{R}_1$ since $\mathbf{R}_2 = \mathbf{0}$. That is, $\mathbf{H} = \mathbf{R}_1^\mathbf{T}\mathbf{Q}_1^\mathbf{T}$. Thus,

$$\mathbf{H}\boldsymbol{\gamma} = \mathbf{H}\mathbf{Q}_2\boldsymbol{\theta} = \mathbf{R}_1^\mathbf{T}\mathbf{Q}_1^\mathbf{T}\mathbf{Q}_2\boldsymbol{\theta} = \mathbf{0}$$

since $\mathbf{Q}_1^\mathbf{T}\mathbf{Q}_2 = \mathbf{0}$. On the other hand, if

$$\mathbf{0} = \mathbf{H}\boldsymbol{\gamma} = \mathbf{R}_1^\mathbf{T}\mathbf{Q}_1^\mathbf{T}\boldsymbol{\gamma},$$

we have $\mathbf{Q}_1^\mathbf{T}\boldsymbol{\gamma} = \mathbf{0}$ since \mathbf{R}_1 is invertible. Thus, $\boldsymbol{\gamma}$ is in the perpendicular subspace of the space spanned by the columns of \mathbf{Q}_1 . That is, $\boldsymbol{\gamma}$ is in the space spanned by the columns of \mathbf{Q}_2 . Thus, there exists a vector $\boldsymbol{\theta}$ such that $\boldsymbol{\gamma} = \mathbf{Q}_2\boldsymbol{\theta}$. These complete the proof. \square

The problem (2.5), is now converted to a conventional penalized regression problem without any constraints:

$$\min_{\boldsymbol{\beta}, \boldsymbol{\theta}} \left\{ \|\mathbf{Y} - \mathbf{Z}\boldsymbol{\beta} - \mathbf{B}\mathbf{Q}_2\boldsymbol{\theta}\|^2 + \lambda(\mathbf{Q}_2\boldsymbol{\theta})^\mathbf{T}\mathbf{P}(\mathbf{Q}_2\boldsymbol{\theta}) \right\}.$$

For a fixed penalty parameter λ , we have

$$\begin{pmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\boldsymbol{\theta}} \end{pmatrix} = \left\{ \begin{pmatrix} \mathbf{Z}^\mathbf{T}\mathbf{Z} & \mathbf{Z}^\mathbf{T}\mathbf{B}\mathbf{Q}_2 \\ \mathbf{Q}_2^\mathbf{T}\mathbf{B}^\mathbf{T}\mathbf{Z} & \mathbf{Q}_2^\mathbf{T}\mathbf{B}^\mathbf{T}\mathbf{B}\mathbf{Q}_2 \end{pmatrix} + \lambda \begin{pmatrix} \mathbf{0} & \\ & \mathbf{Q}_2^\mathbf{T}\mathbf{P}\mathbf{Q}_2 \end{pmatrix} \right\}^{-1} \begin{pmatrix} \mathbf{Z}^\mathbf{T}\mathbf{Y} \\ \mathbf{Q}_2^\mathbf{T}\mathbf{B}^\mathbf{T}\mathbf{Y} \end{pmatrix}.$$

Letting

$$\mathbf{V} = \begin{pmatrix} \mathbf{V}_{11} & \mathbf{V}_{12} \\ \mathbf{V}_{21} & \mathbf{V}_{22} \end{pmatrix} = \begin{pmatrix} \mathbf{Z}^\mathbf{T}\mathbf{Z} & \mathbf{Z}^\mathbf{T}\mathbf{B}\mathbf{Q}_2 \\ \mathbf{Q}_2^\mathbf{T}\mathbf{B}^\mathbf{T}\mathbf{Z} & \mathbf{Q}_2^\mathbf{T}(\mathbf{B}^\mathbf{T}\mathbf{B} + \lambda\mathbf{P})\mathbf{Q}_2 \end{pmatrix}, \quad (2.7)$$

we have

$$\begin{pmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\boldsymbol{\theta}} \end{pmatrix} = \mathbf{V}^{-1} \begin{pmatrix} \mathbf{Z}^\mathbf{T}\mathbf{Y} \\ \mathbf{Q}_2^\mathbf{T}\mathbf{B}^\mathbf{T}\mathbf{Y} \end{pmatrix}.$$

Next let us write

$$\mathbf{V}^{-1} \equiv \mathbf{U} = \begin{pmatrix} \mathbf{U}_{11} & \mathbf{U}_{12} \\ \mathbf{U}_{21} & \mathbf{U}_{22} \end{pmatrix} = \begin{pmatrix} \mathbf{U}_{11} & -\mathbf{U}_{11}\mathbf{V}_{12}\mathbf{V}_{22}^{-1} \\ -\mathbf{U}_{22}\mathbf{V}_{21}\mathbf{V}_{11}^{-1} & \mathbf{U}_{22} \end{pmatrix}, \quad (2.8)$$

where

$$\mathbf{U}_{11}^{-1} = \mathbf{V}_{11} - \mathbf{V}_{12}\mathbf{V}_{22}^{-1}\mathbf{V}_{21} = \mathbf{Z}^\mathbf{T} [\mathbf{I} - \mathbf{B}\mathbf{Q}_2\{\mathbf{Q}_2^\mathbf{T}(\mathbf{B}^\mathbf{T}\mathbf{B} + \lambda\mathbf{P})\mathbf{Q}_2\}^{-1}\mathbf{Q}_2^\mathbf{T}\mathbf{B}^\mathbf{T}] \mathbf{Z}, \quad (2.9)$$

$$\mathbf{U}_{22}^{-1} = \mathbf{V}_{22} - \mathbf{V}_{21}\mathbf{V}_{11}^{-1}\mathbf{V}_{12} = \mathbf{Q}_2^\mathbf{T} [\mathbf{B}^\mathbf{T} \{\mathbf{I} - \mathbf{Z}(\mathbf{Z}^\mathbf{T}\mathbf{Z})^{-1}\mathbf{Z}^\mathbf{T}\} \mathbf{B} + \lambda\mathbf{P}] \mathbf{Q}_2. \quad (2.10)$$

Then the minimizers of (2.7) can be given precisely as follows:

$$\begin{aligned}\hat{\beta} &= \mathbf{U}_{11} \mathbf{Z}^T (\mathbf{I} - \mathbf{B} \mathbf{Q}_2 \mathbf{V}_{22}^{-1} \mathbf{Q}_2^T \mathbf{B}^T) \mathbf{Y} = \mathbf{U}_{11} \mathbf{Z}^T \{ \mathbf{I} - \mathbf{B} \mathbf{Q}_2 \{ \mathbf{Q}_2^T (\mathbf{B}^T \mathbf{B} + \lambda \mathbf{P}) \mathbf{Q}_2 \}^{-1} \mathbf{Q}_2^T \mathbf{B}^T \} \mathbf{Y}, \\ \hat{\theta} &= \mathbf{U}_{22} \mathbf{Q}_2^T \mathbf{B}^T (\mathbf{I} - \mathbf{Z} \mathbf{V}_{11}^{-1} \mathbf{Z}^T) \mathbf{Y} = \mathbf{U}_{22} \mathbf{Q}_2^T \mathbf{B}^T \{ \mathbf{I} - \mathbf{Z} (\mathbf{Z}^T \mathbf{Z})^{-1} \mathbf{Z}^T \} \mathbf{Y}.\end{aligned}$$

Therefore, one obtains the estimators for γ and $g(\cdot)$, respectively:

$$\begin{aligned}\hat{\gamma} &= \mathbf{Q}_2 \hat{\theta} = \mathbf{Q}_2 \mathbf{U}_{22} \mathbf{Q}_2^T \mathbf{B}^T \{ \mathbf{I} - \mathbf{Z} (\mathbf{Z}^T \mathbf{Z})^{-1} \mathbf{Z}^T \} \mathbf{Y}, \\ \hat{g}(\mathbf{x}) &= \mathbf{B}(\mathbf{x})^T \hat{\gamma} = \sum_{\xi \in \mathcal{K}} B_{\xi}(\mathbf{x}) \hat{\gamma}_{\xi}.\end{aligned}\tag{2.11}$$

The fitted values at the n data points are

$$\hat{\mathbf{Y}} = \mathbf{Z} \hat{\beta} + \mathbf{B} \hat{\gamma} = \mathbf{S}(\lambda) \mathbf{Y},$$

where the smoothing or hat matrix is

$$\mathbf{S}(\lambda) = \mathbf{Z} \mathbf{U}_{11} \mathbf{Z}^T (\mathbf{I} - \mathbf{B} \mathbf{Q}_2 \mathbf{V}_{22}^{-1} \mathbf{Q}_2^T \mathbf{B}^T) + \mathbf{B} \mathbf{Q}_2 \mathbf{U}_{22} \mathbf{Q}_2^T \mathbf{B}^T (\mathbf{I} - \mathbf{Z} \mathbf{V}_{11}^{-1} \mathbf{Z}^T).$$

In nonparametric regression, the trace $\text{tr}(\mathbf{S}(\lambda))$ of smoothing matrix $\mathbf{S}(\lambda)$ is often called the degrees of freedom of the model fit. It has the rough interpretation as the equivalent number of parameters and can be thought as a generalization of the definition in linear regression. Finally, we can estimate the variance of the error term, σ^2 by

$$\hat{\sigma}^2 = \frac{\|\mathbf{Y} - \hat{\mathbf{Y}}\|^2}{n - \text{tr}(\mathbf{S}(\lambda))}.\tag{2.12}$$

2.4. Choosing the Triangulation

Triangulation has been extensively investigated in the past few decades, and various packages have been developed. For example, one can use the ‘‘Delaunay’’ algorithm to find a triangulation; see MATLAB program *delaunay.m* or MATHEMATICA function *DelaunayTriangulation*. ‘‘Triangle’’ (Shewchuk, 1996) is also widely used in many applications, and one can download it for free from <http://www.cs.cmu.edu/quake/triangle.html>. It is a C++ program for two-dimensional mesh generation and construction of Delaunay triangulations. ‘‘DistMesh’’ is another method to generate unstructured triangular and tetrahedral meshes; see the *DistMesh* generator on <http://persson.berkeley.edu/distmesh/>. A detailed description of the program is provided by Persson and Strang (2004). We used our own triangulation code in simulation studies and real data analysis below.

As is usual with the one-dimensional (1-D) penalized least squares (PLS) splines, the number of knots is not important given that it is above some minimum depending upon the

degree of the smoothness; see Li and Ruppert (2008). For bivariate PLS splines, Lai and Wang (2013) also observed that the number of triangles N is not very critical, provided N is larger than some threshold. In fact, one of the main advantages of using PLS splines over discrete least squares (DLS) splines is the flexibility of choosing knots in the 1-D setting and choosing triangles in the 2-D setting. For DLS splines, one has to have large enough sample according to the requirement of the degree of splines on each subinterval in the 1-D case or each triangle in the 2-D case to guarantee that a solution can be found. However, there is no such requirement for PLS splines. When the smoothness $r \geq 1$, the only requirement for bivariate PLS splines is that there is at least one triangle containing three points which are not in one line (Lai, 2008). Also, PLS splines perform similar to DLS splines as long as the penalty parameter λ is very small. So in summary, the proposed bivariate PLS splines are very flexible and convenient for data fitting, even for smoothing sparse and unevenly sampled data.

In practice, we recommend that the user first constructs a polygon domain Ω containing all the design points of the data and makes a simple triangulation Δ_0 of Ω by hand or computer, then refines Δ_0 several times to have a triangulation of desired size.

2.5. Penalty Parameter Selection

Selecting a suitable value of smoothing parameter λ is critical to good model fitting. A large value of λ enforces a smoother fitted function with potentially larger fitting errors, while a small value yields a rougher fitted function and potentially smaller fitting errors. Since the in-sample fitting errors can not gauge the prediction property of the fitted function, one should target a criterion function that mimics the out-of-sample performance of the fitted model. The generalized cross-validation (GCV) (Craven and Wahba, 1979; Wahba, 1990) is such a criterion and is widely used for choosing the penalty parameter. We choose the smoothing parameter λ by minimizing the following generalized cross-validation (GCV) criterion

$$\text{GCV}(\lambda) = \frac{n\|\mathbf{Y} - \mathbf{S}(\lambda)\mathbf{Y}\|^2}{\{n - \text{tr}(\mathbf{S}(\lambda))\}^2},$$

over a grid of values of λ . We use the 10-point grid where the values of $\log_{10}(\lambda)$ are equally spaced between -6 and 7 .

3. Asymptotic Results

This section studies the asymptotic properties for the proposed estimators. To discuss these properties, we first introduce some notation. For any function f over the closure of domain Ω , denote $\|f\|_\infty = \sup_{\mathbf{x} \in \Omega} |f(\mathbf{x})|$ the supremum norm of function f and $|f|_{v,\infty} = \max_{i+j=v} \|D_{x_1}^i D_{x_2}^j f(\mathbf{x})\|_\infty$ the maximum norms of all the v th order derivatives of f over Ω .

Let

$$W^{\ell,\infty}(\Omega) = \{f \text{ on } \Omega : |f|_{k,\infty} < \infty, 0 \leq k \leq \ell\} \quad (3.1)$$

be the standard Sobolev space. For any $j = 1, \dots, p$, let z_j be the coordinate mapping that maps \mathbf{z} to its j -th component so that $z_j(\mathbf{Z}_i) = Z_{ij}$, and let

$$h_j = \operatorname{argmin}_{h \in L^2} \|z_j - h\|_{L^2}^2 = \operatorname{argmin}_{h \in L^2} E\{(Z_{ij} - h(\mathbf{X}_i))^2\} \quad (3.2)$$

be the orthogonal projection of z_j onto L^2 .

Before we state the results, we make the following assumptions:

(A1) The random variables Z_{ij} are bounded, uniformly in $i = 1, \dots, n$, $j = 1, \dots, p$.

(A2) The eigenvalues of $E \left\{ \begin{pmatrix} 1 & \mathbf{Z}_i^T \end{pmatrix} \begin{pmatrix} 1 & \mathbf{Z}_i^T \end{pmatrix} \middle| \mathbf{X}_i \right\}$ are bounded away from 0.

(A3) The noise ϵ satisfies that $\lim_{\eta \rightarrow \infty} E[\epsilon^2 I(\epsilon > \eta)] = 0$.

Assumptions (A1)–(A3) are typical in semiparametric smoothing literature, see for instance Huang, Zhang and Zhou (2007) and Wang, et al (2011). The purpose of Assumption (A2) is to ensure that the vector $(1, \mathbf{Z}_i^T)$ is not multicollinear. We next introduce some assumptions on the properties of the true bivariate function in model (1.1) and the bivariate spline space introduced in Section 2.2.

(C1) The bivariate functions $h_j(\cdot)$, $j = 1, \dots, p$, and the true function in model (1.1) $g(\cdot) \in W^{\ell+1,\infty}(\Omega)$ in (3.1) for an integer $\ell \geq 2$.

(C2) For every $s \in \mathbb{S}$ and every $\tau \in \Delta$, there exists a positive constant F_1 , independent of s and τ , such that

$$F_1 \|s\|_{\infty, \tau} \leq \left(\sum_{\mathbf{X}_i \in \tau, i=1, \dots, n} s(\mathbf{X}_i)^2 \right)^{1/2}, \quad \text{for all } \tau \in \Delta. \quad (3.3)$$

(C3) Let F_2 be the largest among the numbers of observations in triangles $\tau \in \Delta$ in the sense that

$$\left(\sum_{\mathbf{X}_i \in \tau, i=1, \dots, n} s(\mathbf{X}_i)^2 \right)^{1/2} \leq F_2 \|s\|_{\infty, \tau}, \quad \text{for all } \tau \in \Delta, \quad (3.4)$$

where $\|s\|_{\infty, \tau}$ denotes the supremum norm of s over triangle τ . The constants F_1 and F_2 defined in (3.3) and (3.4) satisfy $F_2/F_1 = O(1)$.

(C4) The number N of the triangles and the sample size n satisfy that $N = Cn^\gamma$ for some constant $C > 0$ and $1/(\ell + 1) \leq \gamma \leq 1/3$.

(C5) The penalized parameter λ satisfies $\lambda = o(n^{1/2}N^{-1})$.

Condition (C1) describes the requirement for the true bivariate function as usually used in the literature of nonparametric or semiparametric estimation. Condition (C2) ensures the existence of a least squares spline. Although one can get a decent penalized least squares spline fitting with $F_1 = 0$ for some triangles, we need (C2) to study the convergence of bivariate penalized least squares splines. Condition (C3) suggests that we should not put too many observations in one triangle. Condition (C4) requires that the number of triangles is above some minimum depending upon the degree of the spline, which is similar to the requirement of Li and Ruppert (2008) in the univariate case. It also ensures the asymptotic equivalence of the theoretical and empirical inner products/norms defined at the beginning of Section . Condition (C5) is required to reduce the bias of the bivariate spline approximation through “under smoothing” and “choosing smaller λ ”.

To avoid confusion, in the following let β_0 and g_0 be the true parameter value and function in model (1.1). The following theorem states that the rate convergence of $\hat{\beta}$ is root- n and $\hat{\beta}$ is asymptotically normal.

Theorem 1. *Suppose Assumptions (A1)-(A3), (C1)-(C5) hold, then the estimator $\hat{\beta}$ is asymptotically normal, that is,*

$$(n\mathbf{\Sigma})^{1/2}(\hat{\beta} - \beta_0) \rightarrow N(\mathbf{0}, \mathbf{I}),$$

where \mathbf{I} is a $p \times p$ identity matrix,

$$\mathbf{\Sigma} = \sigma^{-2} E\{(\mathbf{Z}_i - \tilde{\mathbf{Z}}_i)(\mathbf{Z}_i - \tilde{\mathbf{Z}}_i)^T\} \quad (3.5)$$

with $\tilde{\mathbf{Z}}_i = \{h_1(\mathbf{X}_i), \dots, h_p(\mathbf{X}_i)\}^T$, for $h_j(\cdot)$ defined in (3.2), $j = 1, \dots, p$. In addition, $\mathbf{\Sigma}$ can be consistently estimated by

$$\mathbf{\Sigma}_n = \frac{1}{n\hat{\sigma}^2} \sum_{i=1}^n (\mathbf{Z}_i - \hat{\mathbf{Z}}_i)(\mathbf{Z}_i - \hat{\mathbf{Z}}_i)^T = \frac{1}{n\hat{\sigma}^2} (\mathbf{Z} - \hat{\mathbf{Z}})^T (\mathbf{Z} - \hat{\mathbf{Z}}). \quad (3.6)$$

where $\hat{\mathbf{Z}}_i$ is the i -th column of $\hat{\mathbf{Z}}^T = \mathbf{Z}^T \mathbf{B} \mathbf{Q}_2 \mathbf{V}_{22}^{-1} \mathbf{Q}_2^T \mathbf{B}^T$ and $\hat{\sigma}^2$ is given by (2.12).

The results in Theorem 1 enable us to construct confidence intervals for the parameters. The next theorem provides the global convergence of the nonparametric estimator $\hat{g}(\cdot)$.

Theorem 2. *Suppose Assumptions (A1)-(A3), (C1)-(C5) hold, then the bivariate penalized estimator $\hat{g}(\cdot)$ in (2.11) is consistent with the true function g_0 , and satisfies that*

$$\|\hat{g} - g_0\|_{L^2} = O_P \left(\frac{\lambda}{n|\Delta|^3} |g_0|_{2,\infty} + \left(1 + \frac{\lambda}{n|\Delta|^5} \right) \frac{F_2}{F_1} |\Delta|^{\ell+1} |g_0|_{\ell,\infty} + \frac{1}{\sqrt{n}|\Delta|} \right).$$

The proofs of the above two theorems are given in Appendix. We notice that the rate of convergence given in Theorem 2 is the same as those for nonparametric spline regression without including the covariate information obtained in Lai and Wang (2013).

4. Simulation

In this section, we carry out two numerical studies to assess the performance of our proposed method. We compare the performance of bivariate penalized splines over triangulations (BPST) with filtered kriging (KRIG), thin plate splines (TPS), soap film smoothing (SOAP) in Wood, Bravington and Hedley (2008), linear finite elements (LFE) and quadratic finite elements (QFE) in Sangalli, Ramsay and Ramsay (2013).

Example 1. In this example, we consider a modified horseshoe domain with the surface test function $g(\cdot)$ used by Wood, Bravington and Hedley (2008) and Sangalli, Ramsay and Ramsay (2013). In particular, for 201×501 grid points over the domain, we simulate data as follows, the response variable Y is generated from the following PLM:

$$Y = \beta_1 Z_1 + \beta_2 Z_2 + g(X_1, X_2) + \epsilon.$$

The random error, ϵ , is generated from an $N(0, \sigma_\epsilon^2)$ distribution with $\sigma_\epsilon = 0.5$. In addition, we set the parameters as $\beta_1 = -1$, $\beta_2 = 1$. For the design of the explanatory variables, Z_1 and Z_2 , two scenarios are considered based on the relationship between the location variables (X_1, X_2) and covariates (Z_1, Z_2) . Under both scenarios, $Z_1 \sim \text{uniform}[-1, 1]$. On the other hand, the variable $Z_2 = \cos[\pi(\rho(X_1^2 + X_2^2) + (1 - \rho)U)]$ where $U \sim \text{uniform}[-1, 1]$ and is independent from (X_1, X_2) as well as Z_1 . We consider both independent design: $\rho = 0.0$ and dependent design: $\rho = 0.7$ in this example. Under both scenarios, 100 Monte Carlo replicates are generated. For each replication, we randomly sample $n = 200$ locations uniformly from the grid points inside the horseshoes domain.

Figure 1 (a) and (b) show the surface and the contour map of the true function $g(\cdot)$, respectively. Figure 1 (c) demonstrates the sampled location points of replicate 1, and Figure 1 (d) and (e) illustrates two different triangulations used in the BPST method. In the first triangulation (\triangle_1), we use 91 triangles and 74 vertices, while in the second one (\triangle_2) we use 109 triangles and 95 vertices.

[Figure 1 about here]

To implement the TPS and KRIG methods, we use the R package *fields* under the standard implementation setting of Furrer, Nychka and Sainand (2011). For KRIG, we try different covariance structures, and we choose the Matérn covariance with smoothness parameter $\nu = 1$, which gives the best prediction. For the SOAP method, we implement it by using R package

mgcv (Wood, Bravington and Hedley, 2008) with 32 interior knots. In addition, a rank 39 (40-knot) cyclic penalized cubic regression spline is used as the boundary curve. For all the methods requiring a smoothing or roughness parameter, GCV is used to choose the values of the parameter.

To see the accuracy of the estimators, we compute the root mean squared error (RMSE) for each of the components based on 100 Monte carlo samples. Table 1 shows the RMSEs of the estimate of the parameters β_1 , β_2 , σ_ε as well as the average mean squared prediction error (MSPE) of the nonlinear function $g(\cdot)$ over all the grid points in the horseshoes shaped domain. From Table 1, one sees that methods SOAP and BPST give very comparable estimates of the parameters β_1 and β_2 , respectively, while BPST method produces the best prediction of the nonlinear function $g(\cdot)$, regardless of the choice of triangulation.

[Table 1 about here]

Figure 2 shows the estimated functions via different methods for the first replicate. For the test function on Figure 2, the BPST estimate looks visually better than the other four estimates. In addition, one sees there is a “leakage effect” in KRIG and TPS estimates and the SOAP, LFE, QFE and BPST methods improve the model fitting of KRIG and TPS. The poor performance of KRIG and TPS is because they do not take the complex boundary into any account and smooth across the gap inappropriately. In addition, from Figure 2, one also sees that the BPST estimators based on \triangle_1 and \triangle_2 are very similar, which agrees our findings for penalized splines that the number of triangles is not very critical for the fitting as long as it is large enough.

[Figure 2 about here]

Next we test the accuracy of the standard error formula in (3.6) for $\hat{\beta}_1$ and $\hat{\beta}_2$, and the results are listed in Table 2. The standard deviations of the estimated parameters are computed based on 100 replications, which can be regarded as the true standard errors (column labeled “SE_{mc}”) and compared with the mean and median of the 100 estimated standard errors calculated using (3.6) (columns labeled “SE_{mean}” and “SE_{median}”, respectively). The column labeled “SE_{mad}” is the interquartile range of the 100 estimated standard errors divided by 1.349, which is a robust estimate of the standard deviation. From Table 2 one observes that the averages or medians of the standard errors calculated using the formula are very close to the true standard deviations, which confirms the accuracy of the proposed standard error formula.

[Table 2 about here]

Finally, for computing time, it takes about 60 seconds to compute BPST in one run of simulation with even 2000 observations over 319 triangles under the computing environment of x64 PC with Intel Dual Core i5. Overall, the proposed algorithm is fast to compute.

Example 2. In this example, we consider a rectangular domain, $[0, 1]^2$, where there is no irregular shape or complex boundaries problem. In this case, classical methods for spatial data analysis, such as KRIG and TPS, will not encounter any difficulty. We obtain the true signal and noisy observation for each coordinate pair lying on a 101×101 -grid over $[0, 1]^2$ using the following model:

$$Y = \mathbf{Z}^T \boldsymbol{\beta} + g(X_1, X_2) + \xi(X_1, X_2) + \epsilon,$$

where $\boldsymbol{\beta} = (-1, 1)^T$ and $g(x_1, x_2) = 5 \sin(2\pi(x_1^2 + x_2^2))$. The random error, ϵ , is generated from an $N(0, \sigma_\epsilon^2)$ distribution with $\sigma_\epsilon = 0.5$, and the process $\xi(\cdot)$ is generated from a stationary gaussian random field with the Matérn(1,1) covariance structure. The components of \mathbf{Z} and ϵ are standard normal, and \mathbf{Z} , ξ and ϵ are independent of each other. Similar to Example 1, we simulate $Z_1 \sim \text{uniform}[-1, 1]$, and $Z_2 = \cos[\pi(\rho(X_1^2 + X_2^2) + (1 - \rho)U)]$, where $\rho = 0.0$ or 0.7 , $U \sim \text{uniform}[-1, 1]$ and is independent from (X_1, X_2) and Z_1 . Next we take 100 Monte Carlo random samples of size $n = 200$ from the 101×101 points.

Figure 4 (a) and (b) display the true sinusoidal surface and the contour map, respectively. 4 (c) represents the Matérn structure. We use the triangulation in Figure 4 (e) and (f), and there are 8 triangles and 9 vertices as well as 32 triangles and 25 vertices, respectively. In addition, the points in Figure 4 (d) demonstrate the sampled location points of replicate 1.

[Figure 4 about here]

Similar to the study in Example 4.1, we also compare the proposed BPST estimator with estimators from the KRIG, TPS, SOAP, LFE and QFE methods, which are implemented in the same way as Example 4.1. Specifically, for KRIG, we choose the true Matérn covariance with smoothness parameter $\nu = 1$. To see the accuracy of the estimators, we compute the RMSEs of the coefficient estimators and the estimator of σ_ϵ . To see the overall prediction accuracy, we make prediction over the 101×101 grid points on the domain for each replication using different methods, and compare the predicted values with the true observations of Y at these grid points, and we report the average mean squared prediction errors (MSPE) over all replications.

All the results are summarized in Table 3. As expected, KRIG works pretty well since the true covariance structure is used in the KRIG fitting. When $\rho = 0$, TPS, KRIG and BPST all perform very well. When $\rho = 0.7$, KRIG and BPST are the best among all the estimators, and both of them outperform the rest of the estimators. In both scenarios, the differences between BPST and KRIG are almost unnoticeable. One also notices that, compared with the

FEM estimators (LEM and QEM), our BPST estimator shows much better performance in terms of both estimation and prediction, because BPST provides a more flexible and easier construction of splines with piecewise polynomials of various degrees and smoothness than the FEM method. Finally as pointed out in Wood, Bravington and Hedley (2008), the standard (linear) FEM method requires a very fine triangulation in order to reach certain approximation power, however, BPST doesn't need such a strict fineness requirement as it uses piecewise polynomials of degree ≥ 5 yielding an higher order approximation power. In fact, we have used 64 times more triangles in the FEM than that for the BPST in our simulation. That is, the BPST is computationally more efficient than the FEM to approximate smooth functions.

[Table 3 about here]

Table 4 lists the accuracy results of the standard error formula in (3.6) for $\hat{\beta}_1$ and $\hat{\beta}_2$ using BPST. From Table 4, one sees that the estimated standard errors based on sample size $n = 200$ are very accurate.

[Table 4 about here]

5. Application to California House Value Data

In this section we apply the proposed method to analyze the California house value data from the StatLib repository. The data appeared in Pace and Barry (1997). The spatial data consists of information of all the block groups in California defined by centroid of census enumeration areas. In the data set, a block group on average includes 1425.5 individuals living in a geographically compact area and there are 20,640 blocks in the final data.

In this paper, we study how different features and factors influence the real estate property prices. The response variable is the median house value (Value). The investigated factors include the median income (MedInc), median house age (Age), the average number of bedrooms (AveBedrms), housing density as reflected by the number of households (Hhd) in each block, and the average occupancy in each household (AveOccup). It is obvious that the location of a house is very crucial for making an accurate prediction, so we also include the latitude (Latitude) and longitude (Longitude) of the block. We model the median house value using the following PLM:

$$\begin{aligned} \log(\text{Value}) = & \beta_0 + \beta_1 \text{MedInc} + \beta_2 \log(\text{Age}) + \beta_3 \log(\text{AveBedrms}) \\ & + \beta_4 \log(\text{AveOccup}) + \beta_5 \log(\text{Hhd}) + g(\text{Latitude}, \text{Longitude}). \end{aligned} \quad (5.1)$$

To fit model (5.1), we use six different methods: KRIG, TPS, SOAP, LFE, QFE, and BPST.

For comparison, we also consider the purely linear model without using the spatial information:

$$\begin{aligned} \log(\text{Value}) = & \beta_0 + \beta_1 \text{MedInc} + \beta_2 \log(\text{Age}) + \beta_3 \log(\text{AveBedrms}) \\ & + \beta_4 \log(\text{AveOccup}) + \beta_5 \log(\text{Hhd}), \end{aligned} \quad (5.2)$$

and fit it using the ordinary linear least squares (OLS) method.

In the following, we report our estimation results for the linear and nonlinear components in model (5.1) by using the OLS, KRIG, TPS, SOAP, LFE, and QFE methods. Table 6 summarizes the coefficient estimation results of all procedures, where the BPST confidence intervals are constructed based on the standard error (s.e.) calculated using (3.6). Note that the coefficient on “log(Age)” is positive 0.165 (*s.e.* = 0.006) in the fitted linear regression model (5.2) via OLS, which is obviously against the common sense in household real estate market. Using the PLM (5.1) with the spatial coordinates, we obtain negative coefficients for “log(Age)” regardless which spatial based method is employed. The coefficient for the variable “log(AveBedrms)” is -0.037 with a standard error 0.018 using OLS. In contrast, all the semiparametric methods unanimously suggest that the coefficient for “log(AveBedrms)” is positive in the PLM in (5.1) after the location effect is controlled. In addition, the coefficient on “log(Hhd)” is positive 0.088 and very significant (*s.e.* = 0.004) in the OLS, but it is statistically insignificant when we apply the PLM (5.1) to the data. In summary, compared with the OLS method, our results from PLM are more consistent with intuitions in the real estate market study. The above counter-intuitive phenomenon in OLS is perhaps due to the model misspecification of (5.2) that completely ignores the location factor of the property.

[Table 6 about here]

Based on the median house values, we classify the houses in the dataset into six different groups: (1) less than 50K, (2) 50K–100K, (3) 100K–150K, (4) 150K–200K, (5) 200K–300K, and (6) greater than 300K, and these groups are plotted in Figure 5. We plot the estimated median house values using different methods in Figure 6, where different colors are used to represent the value of houses as in Figure 5. All the plots in Figures 5 and 6 show that expensive houses are clustered around the major cities and inland house values are lower than coastal house values. The OLS significantly underestimates the coastal enclaves of expensive houses and overestimates the house values in the central valley; see Figure 6 (a). In contrast, the PLM methods in Figure 6 (b)-(g) provide much more accurate estimates.

[Figures 5, 6 about here]

Figure 7 further demonstrates the differences between the estimated house values and the observed house values using seven different methods. As seen in Figure 7, some methods,

especially the OLS, have difficulties in estimating the house values in major metropolitan areas and the central valley. The proposed BPST method dramatically reduces the estimation errors in those areas.

[Figure 7 about here]

To evaluate different methods, we also estimate the out-of-sample prediction errors of each method using the 5-fold cross validation. We randomly split all the observations into five roughly equal-sized parts. For each $k = 1, \dots, 5$, we leave out part k , fit the model to the other four parts (combined), and then obtain predictions for the left-out k th part. Table 5 summarizes the mean squared prediction errors of the logarithm of the median house value based on different methods. As expected, incorporating the spatial information dramatically reduces the prediction errors, and the proposed BPST method provides the most accurate prediction on the median housing values among all the methods.

[Table 5 about here]

6. Concluding Remarks

In this paper, we have considered PLMs for analyzing spatial data. We introduce a framework of bivariate penalized splines over triangulations in the semiparametric estimation. Our work differs from existing works in four major aspects.

First, the proposed estimator solves the problem of “leakage” across the complex domains where many conventional smoothing tools suffer from. The numerical results of the simulation show our method is very effective on both regular and irregular domains.

Secondly, we provide new statistical theories for estimating the PLM for data distributed over complex spatial domains. It is shown that our estimates of both parametric part and non-parametric part of the model enjoy excellent asymptotic properties. In particular, we have shown that our estimates of the coefficients in the parametric part are asymptotically normal and derived the convergence rate of the nonparametric component under regularity conditions. We have also provided a standard error formula for the estimated parameters and our simulation studies show that the standard errors are estimated with good accuracy. The theoretical results provide measures of the effect of covariates after adjusting for the location effect. In addition, they give valuable insights into the accuracy of our estimate of the PLM and permit joint inference for the parameters.

Thirdly, comparing with the conventional FEM, our method provides a more flexible and easier construction of splines with piecewise polynomials of various degrees and various smoothness.

Finally, the proposed method greatly enhances the application of PLMs to spatial data analysis. We don't require the data to be evenly distributed or on regular-spaced grid. When we have regions of sparse data, PLS splines provides a more convenient tool for data fitting than DLS splines since the roughness penalty helps regularize the estimation. Our estimation is computationally fast and efficient since it can formulate a penalized regression problem using a QR decomposition.

This paper leaves open the problem of how to choose a triangulation for a given data set. An optimal triangulation Δ enables us to achieve the best approximation of the given data using a spline space $\mathbb{S}(\Delta)$ of a fixed degree and fixed smoothness. We leave the problem for our future exploration.

Acknowledgment

The first author's research was supported in part by National Science Foundation grants DMS-11-06816 and DMS-15-42332, and the third author's research was supported in part by National Science Foundation grant DMS-15-21537.

Appendices

A.1. Preliminaries

In the following, we use c , C , c_1 , c_2 , C_1 , C_2 , etc. as generic constants, which may be different even in the same line. For functions f_1 and f_2 on $\Omega \times \mathbb{R}^p$, we define the empirical inner product and norm as $\langle f_1, f_2 \rangle_n = \frac{1}{n} \sum_{i=1}^n f_1(\mathbf{X}_i, \mathbf{Z}_i) f_2(\mathbf{X}_i, \mathbf{Z}_i)$ and $\|f_1\|_n^2 = \langle f_1, f_1 \rangle_n$. If f_1 and f_2 are L^2 -integrable, we define the theoretical inner product and theoretical L^2 norm as $\langle f_1, f_2 \rangle_{L^2} = E \{f_1(\mathbf{X}_i, \mathbf{Z}_i) f_2(\mathbf{X}_i, \mathbf{Z}_i)\}$ and $\|f_1\|_{L^2} = \langle f_1, f_1 \rangle_{L^2}$. Furthermore, let $\|\cdot\|_{\mathcal{E}_v}$ be the norm introduced by the inner product $\langle \cdot, \cdot \rangle_{\mathcal{E}_v}$, where

$$\langle g_1, g_2 \rangle_{\mathcal{E}_v} = \int_{\Omega} \left\{ \sum_{i+j=v} \binom{v}{i} (D_{x_1}^i D_{x_2}^j g_1)^2 \right\}^{1/2} \left\{ \sum_{i+j=v} \binom{v}{i} (D_{x_1}^i D_{x_2}^j g_2)^2 \right\}^{1/2} dx_1 dx_2$$

for g_1 and g_2 on Ω .

Lemma A.1. *[Lai and Schumaker (2007)] Let $\{B_{\xi}\}_{\xi \in \mathcal{K}}$ be the Bernstein polynomial basis for spline space \mathbb{S} with smoothness r , where \mathcal{K} stands for an index set. Then there exist positive constants c , C depending on the smoothness r and the shape parameter δ such that*

$$c|\Delta|^2 \sum_{\xi \in \mathcal{K}} |\gamma_{\xi}|^2 \leq \left\| \sum_{\xi \in \mathcal{K}} \gamma_{\xi} B_{\xi} \right\|_{L^2}^2 \leq C|\Delta|^2 \sum_{\xi \in \mathcal{K}} |\gamma_{\xi}|^2$$

for all $\gamma_{\xi}, \xi \in \mathcal{K}$.

With the above stability condition, Lai and Wang (2013) established the following uniform rate at which the empirical inner product approximates the theoretical inner product.

Lemma A.2. [Lemma 2 of the Supplement of Lai and Wang (2013)] Let $g_1 = \sum_{\xi \in \mathcal{K}} c_\xi B_\xi$, $g_2 = \sum_{\zeta \in \mathcal{K}} \tilde{c}_\zeta B_\zeta$ be any spline functions in \mathbb{S} . Under Condition (C4), we have

$$\sup_{g_1, g_2 \in \mathbb{S}} \left| \frac{\langle g_1, g_2 \rangle_n - \langle g_1, g_2 \rangle_{L^2}}{\|g_1\|_{L^2} \|g_2\|_{L^2}} \right| = O_P \left\{ (N \log n)^{1/2} / n^{1/2} \right\}.$$

Lemma A.3. [Corollary of Theorem 6 in Lai (2008)] Assume $g(\cdot)$ is in Sobolev space $W^{\ell+1, \infty}(\Omega)$. For bi-integer (α_1, α_2) with $0 \leq \alpha_1 + \alpha_2 \leq v$, there exists a unique spline fit $g^*(\cdot) \in \mathbb{S}$ such that

$$\|D_{x_1}^{\alpha_1} D_{x_2}^{\alpha_2} (g - g^*)\|_\infty \leq C \frac{F_2}{F_1} |\Delta|^{\ell+1-\alpha_1-\alpha_2} |g|_{\ell+1, \infty},$$

where C is an absolute constant depending on r and δ .

For any smooth bivariate function $g(\cdot)$ and $\lambda > 0$, define

$$s_{\lambda, g} = \operatorname{argmin}_{s \in \mathbb{S}} \sum_{i=1}^n \{g(\mathbf{X}_i) - s(\mathbf{X}_i)\}^2 + \lambda \mathcal{E}_v(s) \quad (\text{A.1})$$

the penalized least squares splines of $g(\cdot)$. Then $s_{0, g}$ is the nonpenalized spline estimator of $g(\cdot)$.

Lemma A.4. Assume $g(\cdot)$ is in Sobolev space $W^{\ell+1, \infty}(\Omega)$. For bi-integer (α_1, α_2) with $0 \leq \alpha_1 + \alpha_2 \leq v$, there exists an absolute constant C depending on r and δ , such that with probability approaching 1,

$$\|D_{x_1}^{\alpha_1} D_{x_2}^{\alpha_2} (g - s_{0, g})\|_\infty \leq C \frac{F_2}{F_1} |\Delta|^{\ell+1-\alpha_1-\alpha_2} |g|_{\ell+1, \infty}.$$

Proof. Note that

$$\begin{aligned} \|D_{x_1}^{\alpha_1} D_{x_2}^{\alpha_2} (g - s_{0, g})\|_\infty &\leq \|D_{x_1}^{\alpha_1} D_{x_2}^{\alpha_2} (g - g^*)\|_\infty + \|D_{x_1}^{\alpha_1} D_{x_2}^{\alpha_2} (g^* - s_{0, g})\|_\infty \\ &\leq \|D_{x_1}^{\alpha_1} D_{x_2}^{\alpha_2} (g - g^*)\|_\infty + \|D_{x_1}^{\alpha_1} D_{x_2}^{\alpha_2} (s_{0, g^*} - g)\|_\infty. \end{aligned}$$

The desired result follows from Lemma A.3, the projection bound result in Golitschek and Schumaker (2002), and the differentiation properties of bivariate splines over triangulations given in Lai and Schumaker (2007). \square

Lemma A.5. Suppose $g(\cdot)$ is in the Sobolev space $W^{\ell+1, \infty}(\Omega)$, and let $s_{\lambda, g}$ be its penalized spline estimator defined in (A.1). Under Conditions (C2) and (C3),

$$\|g - s_{\lambda, g}\|_n = O_P \left\{ \frac{F_2}{F_1} |\Delta|^{\ell+1} |g|_{\ell+1, \infty} + \frac{\lambda}{n |\Delta|^2} \left(|g|_{v, \infty} + \frac{F_2}{F_1} |\Delta|^{\ell+1-v} |g|_{\ell+1, \infty} \right) \right\}.$$

Proof. Note that $s_{\lambda,g}$ is characterized by the orthogonality relations

$$n \langle g - s_{\lambda,g}, u \rangle_n = \lambda \langle s_{\lambda,g}, u \rangle_{\mathcal{E}_v}, \quad \text{for all } u \in \mathbb{S}, \quad (\text{A.2})$$

while $s_{0,g}$ is characterized by

$$\langle g - s_{0,g}, u \rangle_n = 0, \quad \text{for all } u \in \mathbb{S}. \quad (\text{A.3})$$

By (A.2) and (A.3), $n \langle s_{0,g} - s_{\lambda,g}, u \rangle_n = \lambda \langle s_{\lambda,g}, u \rangle_{\mathcal{E}_v}$, for all $u \in \mathbb{S}$. Replacing u by $s_{0,g} - s_{\lambda,g}$ yields that

$$n \|s_{0,g} - s_{\lambda,g}\|_n^2 = \lambda \langle s_{\lambda,g}, s_{0,g} - s_{\lambda,g} \rangle_{\mathcal{E}_v}. \quad (\text{A.4})$$

Thus, by Cauchy-Schwarz inequality,

$$n \|s_{0,g} - s_{\lambda,g}\|_n^2 \leq \lambda \|s_{\lambda,g}\|_{\mathcal{E}_v} \|s_{0,g} - s_{\lambda,g}\|_{\mathcal{E}_v} \leq \lambda \|s_{\lambda,g}\|_{\mathcal{E}_v} \sup_{f \in \mathbb{S}} \left\{ \frac{\|f\|_{\mathcal{E}_v}}{\|f\|_n}, \|f\|_n \neq 0 \right\} \|s_{0,g} - s_{\lambda,g}\|_n.$$

Similarly, using (A.4), we have

$$n \|s_{0,g} - s_{\lambda,g}\|_n^2 = \lambda \left\{ \langle s_{\lambda,g}, s_{0,g} \rangle_{\mathcal{E}_v} - \langle s_{\lambda,g}, s_{\lambda,g} \rangle_{\mathcal{E}_v} \right\} \geq 0.$$

Therefore, by Cauchy-Schwarz inequality,

$$\|s_{\lambda,g}\|_{\mathcal{E}_v}^2 \leq \langle s_{\lambda,g}, s_{0,g} \rangle_{\mathcal{E}_v} \leq \|s_{\lambda,g}\|_{\mathcal{E}_v} \|s_{0,g}\|_{\mathcal{E}_v},$$

which implies that $\|s_{\lambda,g}\|_{\mathcal{E}_v} \leq \|s_{0,g}\|_{\mathcal{E}_v}$. Therefore,

$$\|s_{0,g} - s_{\lambda,g}\|_n \leq n^{-1} \lambda \|s_{0,g}\|_{\mathcal{E}_v} \sup_{f \in \mathbb{S}} \left\{ \frac{\|f\|_{\mathcal{E}_v}}{\|f\|_n}, \|f\|_n \neq 0 \right\}.$$

By Lemma A.4, with probability approaching 1,

$$\begin{aligned} \|s_{0,g}\|_{\mathcal{E}_v} &\leq C_1 A_\Omega \left\{ |g|_{v,\infty} + \sum_{\alpha_1 + \alpha_2 = v} \|D_{x_1}^{\alpha_1} D_{x_2}^{\alpha_2} (g - s_{0,g})\|_\infty \right\} \\ &\leq C_2 A_\Omega \left(|g|_{v,\infty} + \frac{F_2}{F_1} |\Delta|^{\ell+1-v} |g|_{\ell+1,\infty} \right), \end{aligned} \quad (\text{A.5})$$

where A_Ω denotes the area of Ω . By Markov's inequality, for any $f \in \mathbb{S}$, $\|f\|_{\mathcal{E}_v} \leq C |\Delta|^{-2} \|f\|_{L^2}$. Lemma (A.2) implies that $\sup_{f \in \mathbb{S}} \{ \|f\|_n / \|f\|_{L^2} \} \geq 1 - O_P \{ (N \log n)^{1/2} / n^{1/2} \}$. Thus, we have

$$\sup_{f \in \mathbb{S}} \left\{ \frac{\|f\|_{\mathcal{E}_v}}{\|f\|_n}, \|f\|_n \neq 0 \right\} \leq C |\Delta|^{-2} \left[1 - O_P \left\{ (N \log n)^{1/2} / n^{1/2} \right\} \right]^{-1/2} = O_P \left(|\Delta|^{-2} \right). \quad (\text{A.6})$$

Therefore,

$$\|s_{0,g} - s_{\lambda,g}\|_n = O_P \left\{ \frac{\lambda}{n|\Delta|^2} \left(|g|_{v,\infty} + \frac{F_2}{F_1} |\Delta|^{\ell+1-v} |g|_{\ell+1,\infty} \right) \right\},$$

and

$$\|g - s_{\lambda,g}\|_n \leq \|g - s_{0,g}\|_n + \|s_{0,g} - s_{\lambda,g}\|_n.$$

By Lemma A.4,

$$\|g - s_{0,g}\|_n \leq \|g - s_{0,g}\|_\infty = O_P \left(\frac{F_2}{F_1} |\Delta|^{\ell+1} |g|_{\ell+1,\infty} \right).$$

Thus, the desired result is established. \square

Lemma A.6. *Under Assumptions (A1), (A2), (C4), (C5), there exist constants $0 < c_U < C_U < \infty$, such that with probability approaching 1 as $n \rightarrow \infty$, $c_U \mathbf{I}_{p \times p} \leq n \mathbf{U}_{11} \leq C_U \mathbf{I}_{p \times p}$, where \mathbf{U}_{11} is given in (2.8).*

Proof. Denote by

$$\mathbf{\Gamma}_\lambda = \frac{1}{n} (\mathbf{B}^T \mathbf{B} + \lambda \mathbf{P}) = \left[\frac{1}{n} \sum_{i=1}^n B_\xi(\mathbf{X}_i) B_\zeta(\mathbf{X}_i) + \frac{\lambda}{n} \langle B_\xi, B_\zeta \rangle_{\mathcal{E}_v} \right]_{\xi, \zeta \in \mathcal{K}}$$

a symmetric positive definite matrix. Then for \mathbf{V}_{22} defined in (2.7), we can rewrite it as $\mathbf{V}_{22} = n \mathbf{Q}_2^T \mathbf{\Gamma}_\lambda \mathbf{Q}_2$. Let $\alpha_{\min}(\lambda)$ and $\alpha_{\max}(\lambda)$ be the smallest and largest eigenvalues of $\mathbf{\Gamma}_\lambda$. As shown in the proof of Theorem 2 in the Supplement of Lai and Wang (2013), there exist positive constants $0 < c_3 < C_3$ such that under Conditions (C4) and (C5), with probability approaching 1, we have

$$c_3 |\Delta|^2 \leq \alpha_{\min}(\lambda) \leq \alpha_{\max}(\lambda) \leq C_3 \left(|\Delta|^2 + \frac{\lambda}{n|\Delta|^2} \right).$$

Therefore, we have

$$c_4 \left(|\Delta|^2 + \frac{\lambda}{n|\Delta|^2} \right)^{-1} \|\mathbf{a}\|^2 \leq n \mathbf{a}^T \mathbf{V}_{22}^{-1} \mathbf{a} = \mathbf{a}^T (\mathbf{Q}_2^T \mathbf{\Gamma}_\lambda \mathbf{Q}_2)^{-1} \mathbf{a} \leq C_4 |\Delta|^{-2} \|\mathbf{a}\|^2.$$

Thus, by Assumption (A2), we have with probability approaching 1

$$c_5 \left(|\Delta|^2 + \frac{\lambda}{n|\Delta|^2} \right)^{-1} |\Delta|^2 \|\mathbf{a}\|^2 \leq \mathbf{a}^T \mathbf{V}_{12} \mathbf{V}_{22}^{-1} \mathbf{V}_{21} \mathbf{a} = \mathbf{a}^T \mathbf{Z}^T \mathbf{B} \mathbf{Q}_2 \mathbf{V}_{22}^{-1} \mathbf{Q}_2^T \mathbf{B}^T \mathbf{Z} \mathbf{a} \leq C_5 \|\mathbf{a}\|^2. \quad (\text{A.7})$$

According to (2.8) and (2.9), we have

$$(n \mathbf{U}_{11})^{-1} = n^{-1} (\mathbf{V}_{11} - \mathbf{V}_{12} \mathbf{V}_{22}^{-1} \mathbf{V}_{21}) = n^{-1} (\mathbf{Z}^T \mathbf{Z} - \mathbf{V}_{12} \mathbf{V}_{22}^{-1} \mathbf{V}_{21}).$$

The desired result follows from Assumptions (A1), (A2) and (A.7). \square

A.2. Proof of Theorem 1

Let $\mu_i = \mathbf{Z}_i^T \beta_0 + g_0(\mathbf{X}_i)$, $\boldsymbol{\mu}^T = (\mu_1, \dots, \mu_n)$, and let $\boldsymbol{\epsilon}^T = (\epsilon_1, \dots, \epsilon_n)$. Define

$$\tilde{\beta}_\mu = \mathbf{U}_{11} \mathbf{Z}^T (\mathbf{I} - \mathbf{B} \mathbf{Q}_2 \mathbf{V}_{22}^{-1} \mathbf{Q}_2^T \mathbf{B}^T) \boldsymbol{\mu}, \quad (\text{A.8})$$

$$\tilde{\beta}_\epsilon = \mathbf{U}_{11} \mathbf{Z}^T (\mathbf{I} - \mathbf{B} \mathbf{Q}_2 \mathbf{V}_{22}^{-1} \mathbf{Q}_2^T \mathbf{B}^T) \boldsymbol{\epsilon}. \quad (\text{A.9})$$

Then $\hat{\beta} - \beta_0 = (\tilde{\beta}_\mu - \beta_0) + \tilde{\beta}_\epsilon$.

Lemma A.7. *Under Assumptions (A1), (A2), (C1)-(C5), $\|\tilde{\beta}_\mu - \beta_0\| = o_P(n^{-1/2})$ for $\tilde{\beta}_\mu$ in (A.8).*

Proof. Let $\mathbf{g}_0 = (g_0(\mathbf{X}_1), \dots, g_0(\mathbf{X}_n))^T$. It is clear that

$$\begin{aligned} \tilde{\beta}_\mu - \beta_0 &= \mathbf{U}_{11} \mathbf{Z}^T (\mathbf{I} - \mathbf{B} \mathbf{Q}_2 \mathbf{V}_{22}^{-1} \mathbf{Q}_2^T \mathbf{B}^T) \mathbf{g}_0 \\ &= \mathbf{U}_{11} \mathbf{Z}^T [\mathbf{g}_0 - \mathbf{B} \mathbf{Q}_2 \{\mathbf{Q}_2^T (\mathbf{B}^T \mathbf{B} + \lambda \mathbf{P}) \mathbf{Q}_2\}^{-1} \mathbf{Q}_2^T \mathbf{B}^T \mathbf{g}_0] \\ &= n \mathbf{U}_{11} \mathbf{A}, \end{aligned}$$

where $\mathbf{A} = (A_1, \dots, A_p)^T$, with

$$A_j = n^{-1} \mathbf{Z}_j^T [\mathbf{g}_0 - \mathbf{B} \mathbf{Q}_2 \{\mathbf{Q}_2^T (\mathbf{B}^T \mathbf{B} + \lambda \mathbf{P}) \mathbf{Q}_2\}^{-1} \mathbf{Q}_2^T \mathbf{B}^T \mathbf{g}_0]$$

for $\mathbf{Z}_j^T = (Z_{1j}, \dots, Z_{nj})$. Next we derive the order of A_j , $1 \leq j \leq p$, as follows. For any $g_j \in \mathbb{S}$, by (A.2) we have

$$A_j = \langle z_j, g_0 - s_{\lambda, g_0} \rangle_n = \langle z_j - g_j, g_0 - s_{\lambda, g_0} \rangle_n + \frac{\lambda}{n} \langle s_{\lambda, g_0}, g_j \rangle_{\mathcal{E}_v}.$$

For any $j = 1, \dots, p$, let $h_j(\cdot)$ be the function $h(\cdot)$ that minimizes $E\{Z_{ij} - h(\mathbf{X}_i)\}^2$ as defined in (3.2). According to Lemma A.3, there exists a function $\tilde{h}_j \in \mathbb{S}$ satisfy

$$\|\tilde{h}_j - h_j\|_\infty \leq C \frac{F_2}{F_1} |\Delta|^{\ell+1} |h_j|_{\ell+1, \infty}, \quad (\text{A.10})$$

then

$$A_j = \langle z_j - h_j, g_0 - s_{\lambda, g_0} \rangle_n + \langle h_j - \tilde{h}_j, g_0 - s_{\lambda, g_0} \rangle_n + \frac{\lambda}{n} \langle s_{\lambda, g_0}, \tilde{h}_j \rangle_{\mathcal{E}_v} = A_{j,1} + A_{j,2} + A_{j,3}.$$

Since h_j satisfies $\langle z_j - h_j, \psi \rangle_{L_2(\Omega)} = 0$ for any $\psi \in L_2(\Omega)$, $E(A_{j,1}) = 0$. According to Proposition 1 in Lai and Wang (2013),

$$\|g_0 - s_{\lambda, g_0}\|_\infty = O_P \left\{ \frac{F_2}{F_1} |\Delta|^{\ell+1} |g_0|_{\ell+1, \infty} + \frac{\lambda}{n |\Delta|^3} \left(|g_0|_{2, \infty} + \frac{F_2}{F_1} |\Delta|^{\ell-1} |g_0|_{\ell+1, \infty} \right) \right\}.$$

Next,

$$\text{Var}(A_{j,1}) = \frac{1}{n^2} \sum_{i=1}^n E[\{Z_{ij} - h_j(\mathbf{X}_i)\} (g_0 - s_{\lambda,g_0})]^2 \leq \frac{\|g_0 - s_{\lambda,g_0}\|_\infty^2}{n} \|z_j - h_j\|_{L^2}^2,$$

which together with $E(A_{j,1}) = 0$ implies that

$$|A_{j,1}| = O_P \left\{ \frac{F_2}{n^{1/2} F_1} |\Delta|^{\ell+1} |g_0|_{\ell+1,\infty} + \frac{\lambda}{n^{3/2} |\Delta|^3} \left(|g_0|_{2,\infty} + \frac{F_2}{F_1} |\Delta|^{\ell-1} |g_0|_{\ell+1,\infty} \right) \right\}.$$

Cauchy-Schwartz inequality, Lemma A.5 and (A.10) imply that

$$\begin{aligned} |A_{j,2}| &\leq \|h_j - \tilde{h}_j\|_n \|g_0 - s_{\lambda,g_0}\|_n \\ &= O_P \left(\frac{F_2}{F_1} |\Delta|^{\ell+1} |h_j|_{\ell+1,\infty} \right) \\ &\times O_P \left\{ \frac{F_2}{F_1} |\Delta|^{\ell+1} |g_0|_{\ell+1,\infty} + \frac{\lambda}{n |\Delta|^2} \left(|g_0|_{2,\infty} + \frac{F_2}{F_1} |\Delta|^{\ell-1} |g_0|_{\ell+1,\infty} \right) \right\}. \end{aligned}$$

Finally, by (A.5), we have

$$\begin{aligned} |A_{j,3}| &\leq \frac{\lambda}{n} \|s_{\lambda,g_0}\|_{\mathcal{E}_v} \|\tilde{h}_j\|_{\mathcal{E}_v} \leq \frac{\lambda}{n} \|s_{0,g_0}\|_{\mathcal{E}_v} \|\tilde{h}_j\|_{\mathcal{E}_v} \\ &\leq \frac{\lambda}{n} C_1 \left(|g_0|_{2,\infty} + \frac{F_2}{F_1} |\Delta|^{\ell-1} |g_0|_{\ell+1,\infty} \right) \left(|h_j|_{2,\infty} + \frac{F_2}{F_1} |\Delta|^{\ell-1} |h_j|_{\ell+1,\infty} \right). \end{aligned}$$

Combining all the above results yields that

$$|A_j| = O_P \left[n^{-1/2} \left\{ \frac{F_2}{F_1} |\Delta|^{\ell+1} |g_0|_{\ell+1,\infty} + \frac{\lambda}{n |\Delta|^3} \left(|g_0|_{2,\infty} + \frac{F_2}{F_1} |\Delta|^{\ell-1} |g_0|_{\ell+1,\infty} \right) \right\} \right]$$

for $j = 1, \dots, p$. By Assumptions (C3)-(C5), $|A_j| = o_P(n^{-1/2})$, for $j = 1, \dots, p$. In addition, we have $nU_{11} = O_P(1)$ according to Lemma A.6. Therefore, $\|\tilde{\beta}_\mu - \beta_0\| = o_P(n^{-1/2})$. \square

Lemma A.8. *Under Assumptions (A1)-(A3) and (C1)-(C5), as $n \rightarrow \infty$,*

$$\left[\text{Var} \left(\tilde{\beta}_\epsilon | \{(\mathbf{X}_i, \mathbf{Z}_i), i = 1, \dots, n\} \right) \right]^{-1/2} \tilde{\beta}_\epsilon \longrightarrow N(0, \mathbf{I}_{p \times p}),$$

where $\tilde{\beta}_\epsilon$ is given in (A.9).

Proof. Note that

$$\tilde{\beta}_\epsilon = \mathbf{U}_{11} \mathbf{Z}^T (\mathbf{I} - \mathbf{B} \mathbf{Q}_2 \mathbf{V}_{22}^{-1} \mathbf{Q}_2^T \mathbf{B}^T) \epsilon.$$

For any $\mathbf{b} \in \mathbb{R}^p$ with $\|\mathbf{b}\| = 1$, we can write $\mathbf{b}^T \tilde{\beta}_\epsilon = \sum_{i=1}^n \alpha_i \epsilon_i$, where

$$\alpha_i^2 = n^{-2} \mathbf{b}^T (n \mathbf{U}_{11}) (\mathbf{Z}_i^T - \mathbf{V}_{12} \mathbf{V}_{22}^{-1} \mathbf{Q}_2^T \mathbf{B}_i) (\mathbf{Z}_i - \mathbf{B}_i^T \mathbf{Q}_2 \mathbf{V}_{22}^{-1} \mathbf{V}_{21}) (n \mathbf{U}_{11}) \mathbf{b},$$

and conditioning on $\{(\mathbf{X}_i, \mathbf{Z}_i), i = 1, \dots, n\}$, $\alpha_i \epsilon_i$'s are independent. By Lemma A.6, we have

$$\max_{1 \leq i \leq n} \alpha_i^2 \leq Cn^{-2} \max_{1 \leq i \leq n} \left\{ \|\mathbf{Z}_i\|^2 + \|\mathbf{V}_{12} \mathbf{V}_{22}^{-1} \mathbf{Q}_2^T \mathbf{B}_i\|^2 \right\},$$

where for any $\mathbf{a} \in \mathbb{R}^p$,

$$\mathbf{a}^T \mathbf{V}_{12} \mathbf{V}_{22}^{-1} \mathbf{Q}_2^T \mathbf{B}_i \mathbf{a} = n^{-1} \mathbf{a}^T \mathbf{V}_{12} (\mathbf{Q}_2^T \mathbf{\Gamma}_\lambda \mathbf{Q}_2)^{-1} \mathbf{Q}_2^T \mathbf{B}_i \mathbf{a} \leq Cn^{-1} |\Delta|^{-2} \mathbf{a}^T \mathbf{Z}^T \mathbf{B} \mathbf{B}_i \mathbf{a},$$

and the j -th component of $n^{-1} \mathbf{Z}^T \mathbf{B} \mathbf{B}_i$ is $\frac{1}{n} \sum_{i'=1}^n Z_{i'j} \sum_{\xi \in \mathcal{K}} B_\xi(\mathbf{X}_{i'}) B_\xi(\mathbf{X}_i)$. Using Assumptions (A1) and (A2), we have $E \left\{ \frac{1}{n} \sum_{i'=1}^n Z_{i'j} \sum_{\xi \in \mathcal{K}} B_\xi(\mathbf{X}_{i'}) B_\xi(\mathbf{X}_i) \right\}^2 = O(1)$, for large n , thus with probability approaching 1,

$$\max_{1 \leq i \leq n} \left| \frac{1}{n} \sum_{i'=1}^n \sum_{\xi \in \mathcal{K}} Z_{i'j} B_\xi(\mathbf{X}_{i'}) B_\xi(\mathbf{X}_i) \right| = O_P(1), \quad \max_{1 \leq i \leq n} \|\mathbf{V}_{12} \mathbf{V}_{22}^{-1} \mathbf{Q}_2^T \mathbf{B}_i\|^2 = O_P(|\Delta|^{-2}).$$

Therefore, $\max_{1 \leq i \leq n} \alpha_i^2 = O_P(n^{-2} |\Delta|^{-2})$. Next, with probability approaching 1,

$$\begin{aligned} \sum_{i=1}^n \alpha_i^2 &= \text{Var} \left[\mathbf{b}^T \tilde{\boldsymbol{\beta}}_\epsilon | \{(\mathbf{X}_i, \mathbf{Z}_i), i = 1, \dots, n\} \right] \\ &= \mathbf{b}^T \mathbf{U}_{11} \mathbf{Z}^T (\mathbf{I} - \mathbf{B} \mathbf{Q}_2 \mathbf{V}_{22}^{-1} \mathbf{Q}_2^T \mathbf{B}^T) (\mathbf{I} - \mathbf{B} \mathbf{Q}_2 \mathbf{V}_{22}^{-1} \mathbf{Q}_2^T \mathbf{B}^T) \mathbf{Z} \mathbf{U}_{11} \mathbf{b} \sigma^2 \\ &= n^{-1} \mathbf{b}^T (n \mathbf{U}_{11}) \left\{ n^{-1} \sum_{i=1}^n (\mathbf{Z}_i - \hat{\mathbf{Z}}_i)(\mathbf{Z}_i - \hat{\mathbf{Z}}_i)^T \right\} (n \mathbf{U}_{11}) \mathbf{b} \sigma^2, \end{aligned} \quad (\text{A.11})$$

where $\hat{\mathbf{Z}}_i$ is the i -th column of $\mathbf{Z}^T \mathbf{B} \mathbf{Q}_2 \mathbf{V}_{22}^{-1} \mathbf{Q}_2^T \mathbf{B}^T$. Using Lemma A.6 again, we have $\sum_{i=1}^n \alpha_i^2 \geq cn^{-1}$. So $\max_{1 \leq i \leq n} \alpha_i^2 / \sum_{i=1}^n \alpha_i^2 = O_P(n^{-1} |\Delta|^{-2}) = o_P(1)$ from Assumption (C4). By Linderberg-Feller CLT, we have $\sum_{i=1}^n \alpha_i \epsilon_i / (\sum_{i=1}^n \alpha_i^2)^{1/2} \rightarrow N(0, 1)$. Then the desired result follows. \square

For any $j = 1, \dots, p$ and $\lambda > 0$, define

$$s_{\lambda, z_j} = \arg \min_{s \in \mathbb{S}} \sum_{i=1}^n \{z_j(\mathbf{X}_i) - s(\mathbf{X}_i)\}^2 + \lambda \mathcal{E}_v(s), \quad (\text{A.12})$$

where z_j is the coordinate mapping that maps \mathbf{z} to its j -th component.

Lemma A.9. *Under Assumptions (A1), (A2), (C2), (C3), for s_{λ, z_j} defined in (A.12), $\|s_{0, z_j} - s_{\lambda, z_j}\|_n = O_P(\lambda n^{-1} |\Delta|^{-5})$, $j = 1, \dots, p$.*

Proof. Note that

$$n \langle z_j - s_{\lambda, z_j}, u \rangle_n = \lambda \langle s_{\lambda, z_j}, u \rangle_{\mathcal{E}_v}, \quad \langle z_j - s_{0, z_j}, u \rangle_n = 0 \quad \text{for all } u \in \mathcal{S},$$

Inserting $u = s_{0,z_j} - s_{\lambda,z_j}$ in the above yields that

$$n \|s_{0,z_j} - s_{\lambda,z_j}\|_n^2 = \lambda \langle s_{\lambda,z_j}, s_{0,z_j} - s_{\lambda,z_j} \rangle_{\mathcal{E}_v} = \lambda (\langle s_{\lambda,z_j}, s_{0,z_j} \rangle_{\mathcal{E}_v} - \langle s_{\lambda,z_j}, s_{\lambda,z_j} \rangle_{\mathcal{E}_v}).$$

By Cauchy-Schwarz inequality, $\|s_{\lambda,z_j}\|_{\mathcal{E}_v}^2 \leq \langle s_{\lambda,z_j}, s_{0,z_j} \rangle_{\mathcal{E}_v} \leq \|s_{\lambda,z_j}\|_{\mathcal{E}_v} \|s_{0,z_j}\|_{\mathcal{E}_v}$, which implies

$$\|s_{\lambda,z_j}\|_{\mathcal{E}_v} \leq \|s_{0,z_j}\|_{\mathcal{E}_v}. \quad (\text{A.13})$$

By (A.6), we have for large n

$$n \|s_{0,z_j} - s_{\lambda,z_j}\|_n^2 \leq \lambda \|s_{\lambda,z_j}\|_{\mathcal{E}_v} \|s_{0,z_j} - s_{\lambda,z_j}\|_n \times O_P(|\Delta|^{-2}),$$

thus, $\|s_{0,z_j} - s_{\lambda,z_j}\|_n \leq \|s_{0,z_j}\|_{\mathcal{E}_v} \times O_P(\lambda n^{-1} |\Delta|^{-2})$. Markov's inequality implies that

$$\|s_{0,z_j}\|_{\mathcal{E}_v} \leq \frac{C_1}{|\Delta|^2} \|s_{0,z_j}\|_{\infty}. \quad (\text{A.14})$$

Note that $\|s_{0,z_j}\|_{\infty} \leq C |\Delta|^{-2} \max_{\xi \in \mathcal{K}} |n^{-1} \sum_{i=1}^n B_{\xi}(\mathbf{X}_i) Z_{ij}|$ with probability approaching one. According to Assumptions (A1) and (A2), $\max_{\xi \in \mathcal{K}} |n^{-1} \sum_{i=1}^n B_{\xi}(\mathbf{X}_i) Z_{ij}| = O_P(|\Delta|)$. The desired results follows. \square

Lemma A.10. *Under Assumptions (A1)-(A3) and (C1)-(C5), for the covariance matrix Σ defined in (3.5), we have $c_{\Sigma}^* \mathbf{I}_p \leq \Sigma \leq C_{\Sigma}^* \mathbf{I}_p$, and $\text{Var}(\tilde{\beta}_{\epsilon} | \{(\mathbf{X}_i, \mathbf{Z}_i), i = 1, \dots, n\}) = n^{-1} \Sigma + o_P(1)$.*

Proof. According to (A.11),

$$\text{Var}(\tilde{\beta}_{\epsilon} | \{(\mathbf{X}_i, \mathbf{Z}_i), i = 1, \dots, n\}) = n^{-1} (n \mathbf{U}_{11}) \left\{ n^{-1} \sum_{i=1}^n (\mathbf{Z}_i - \hat{\mathbf{Z}}_i)(\mathbf{Z}_i - \hat{\mathbf{Z}}_i)^T \right\} (n \mathbf{U}_{11}) \sigma^2.$$

By the definition of \mathbf{U}_{11}^{-1} in (2.9), we have

$$(n \mathbf{U}_{11})^{-1} = \frac{1}{n} \sum_{i=1}^n \mathbf{Z}_i (\mathbf{Z}_i - \hat{\mathbf{Z}}_i)^T = \left(\langle z_j, z_{j'} - s_{\lambda,z_{j'}} \rangle_n \right)_{1 \leq j, j' \leq p}.$$

As in the proof of Lemma A.7, let $\tilde{h}_j \in \mathbb{S}$ and h_j satisfy (A.10). Then we have

$$\langle z_j, z_{j'} - s_{\lambda,z_{j'}} \rangle_n = \langle z_j - \tilde{h}_j, z_{j'} - s_{\lambda,z_{j'}} \rangle_n + \frac{\lambda}{n} \langle s_{\lambda,z_{j'}}, \tilde{h}_j \rangle_{\mathcal{E}_v}. \quad (\text{A.15})$$

By (A.5), (A.13) and (A.14), we have

$$\begin{aligned} \left| \langle s_{\lambda,z_{j'}}, \tilde{h}_{j'} \rangle_{\mathcal{E}_v} \right| &\leq \|s_{\lambda,z_{j'}}\|_{\mathcal{E}_v} \|\tilde{h}_{j'}\|_{\mathcal{E}_v} \leq \|s_{0,z_{j'}}\|_{\mathcal{E}_v} \|\tilde{h}_{j'}\|_{\mathcal{E}_v} \\ &\leq \frac{C}{|\Delta|^2} \|s_{0,z_{j'}}\|_{\infty} \|\tilde{h}_{j'}\|_{\mathcal{E}_v} \leq \frac{CC^*}{|\Delta|^3} \left(|h'_{j'}|_{2,\infty} + \frac{F_2}{F_1} |\Delta|^{\ell+1-v} |h'_{j'}|_{\ell+1,\infty} \right). \end{aligned}$$

Note that

$$\begin{aligned} \langle z_j - \tilde{h}_j, z_{j'} - s_{\lambda, z_{j'}} \rangle_n &= \langle z_j - h_j, z_{j'} - h_{j'} \rangle_n + \langle h_j - \tilde{h}_j, h_{j'} - \tilde{h}_{j'} \rangle_n + \langle z_j - h_j, h_{j'} - \tilde{h}_{j'} \rangle_n \\ &\quad + \langle h_j - \tilde{h}_j, z_{j'} - h_{j'} \rangle_n + \langle z_j - h_j, \tilde{h}_{j'} - s_{\lambda, z_{j'}} \rangle_n + \langle h_j - \tilde{h}_j, \tilde{h}_{j'} - s_{\lambda, z_{j'}} \rangle_n. \end{aligned} \quad (\text{A.16})$$

According to (A.10), the second term on the right side of (A.16) satisfies that

$$\left| \langle h_j - \tilde{h}_j, h_{j'} - \tilde{h}_{j'} \rangle_\infty \right| \leq \|h_j - \tilde{h}_j\|_\infty \|h_{j'} - \tilde{h}_{j'}\|_\infty = o_P(1).$$

By Lemma A.2 and (A.10), the third term on the right side of (A.16) satisfies that

$$\left| \langle z_j - h_j, h_{j'} - \tilde{h}_{j'} \rangle_n \right| \leq \{\|z_j - h_j\|_{L^2}(1 + o_P(1))\} \|h_{j'} - \tilde{h}_{j'}\|_\infty = o_P(1).$$

Similarly, we have

$$\left| \langle h_j - \tilde{h}_j, z_{j'} - h_{j'} \rangle_n \right| = o_P(1).$$

From the triangle inequality, we have

$$\|\tilde{h}_j - s_{\lambda, z_j}\|_n \leq \|\tilde{h}_j - h_j\|_n + \|h_j - s_{0, z_j}\|_n + \|s_{0, z_j} - s_{\lambda, z_j}\|_n.$$

According to (A.10) and Lemma A.9, we have

$$\|\tilde{h}_j - s_{\lambda, z_j}\|_n \leq \|h_j - s_{0, z_j}\|_n + o_P(1).$$

Define $h_{j,n}^* = \operatorname{argmin}_{h \in \mathbb{S}} \|z_j - h\|_{L^2}$, then, from the triangle inequality, we have

$$\|h_j - s_{0, z_j}\|_n \leq \|h_j - h_{j,n}^*\|_n + \|h_{j,n}^* - s_{0, z_j}\|_n$$

Note that $\|h_j - h_{j,n}^*\|_{L^2} = o_P(1)$. Lemma A.2 implies that $\|h_j - h_{j,n}^*\|_n = o_P(1)$. Next note that $\|s_{0, z_j} - h_{j,n}^*\|_{L^2}^2 = \|z_j - s_{0, z_j}\|_{L^2}^2 - \|z_j - h_{j,n}^*\|_{L^2}^2$ and $\|z_j - s_{0, z_j}\|_n \leq \|z_j - h_{j,n}^*\|_n$. Using Lemma A.2 again, we have

$$\|s_{0, z_j} - h_{j,n}^*\|_{L^2}^2 = o_P(\|z_j - h_{j,n}^*\|_{L^2}^2) + o_P(\|z_j - s_{0, z_j}\|_{L^2}^2).$$

Since there exists a constant C such that $\|z_j - h_{j,n}^*\|_{L^2} \leq C$, so we have

$$\|z_j - s_{0, z_j}\|_{L^2} \leq \|z_j - h_{j,n}^*\|_{L^2} + \|h_{j,n}^* - s_{0, z_j}\|_{L^2} \leq C + \|h_{j,n}^* - s_{0, z_j}\|_{L^2}.$$

Therefore, $\|h_{j,n}^* - s_{0, z_j}\|_{L^2} = o_P(1)$. Lemma A.2 implies that $\|h_{j,n}^* - s_{0, z_j}\|_n = o_P(1)$. As a consequence,

$$\|s_{0, z_j} - h_j\|_n = o_P(1). \quad (\text{A.17})$$

For the fifth item, by Lemma A.2 and (A.17), we have

$$\left| \langle z_j - h_j, \tilde{h}_{j'} - s_{\lambda, z_{j'}} \rangle_n \right| \leq \{\|z_j - h_j\|_{L^2}(1 + o_P(1))\} \{\|h_j - s_{0, z_j}\|_n + o_P(1)\} = o_P(1).$$

Similarly, for the sixth item, we have

$$\left| \langle h_j - \tilde{h}_j, \tilde{h}_{j'} - s_{\lambda, z_{j'}} \rangle_n \right| \leq \|h_j - \tilde{h}_j\|_n \{ \|h_j - s_{0, z_j}\|_n + o_P(1) \} = o_P(1). \quad (\text{A.18})$$

Combining the above results from (A.15) to (A.18) gives that

$$\langle z_j, z_{j'} - s_{\lambda, z_{j'}} \rangle_n = \langle z_j - h_j, z_{j'} - h_{j'}^* \rangle_n + o_P(1).$$

Therefore,

$$(n\mathbf{U}_{11})^{-1} = \frac{1}{n} \sum_{i=1}^n (\mathbf{Z}_i - \tilde{\mathbf{Z}}_i)(\mathbf{Z}_i - \tilde{\mathbf{Z}}_i)^T + o_P(1) = E[(\mathbf{Z}_i - \tilde{\mathbf{Z}}_i)(\mathbf{Z}_i - \tilde{\mathbf{Z}}_i)^T] + o_P(1).$$

Hence,

$$\text{Var} \left(\tilde{\beta}_\epsilon \mid \{(\mathbf{X}_i, \mathbf{Z}_i), i = 1, \dots, n\} \right) = n^{-1} \boldsymbol{\Sigma}^{-1} + o_P(1). \quad \square$$

A.3. Proof of Theorem 2

Let $\mathbf{H}_Z = \mathbf{I} - \mathbf{Z}(\mathbf{Z}^T \mathbf{Z})^{-1} \mathbf{Z}^T$, then

$$\hat{\boldsymbol{\theta}} = \mathbf{U}_{22} \mathbf{Q}_2^T \mathbf{B}^T \mathbf{H}_Z \mathbf{Y} = \mathbf{U}_{22} \mathbf{Q}_2^T \mathbf{B}^T \mathbf{H}_Z \mathbf{g}_0 + \mathbf{U}_{22} \mathbf{Q}_2^T \mathbf{B}^T \mathbf{H}_Z \boldsymbol{\epsilon} = \tilde{\boldsymbol{\theta}}_\mu + \tilde{\boldsymbol{\theta}}_\epsilon.$$

According to Lemma A.3, $\|s_{0, g_0} - g_0\|_\infty \leq C \frac{F_2}{F_1} |\Delta|^{\ell+1} |g_0|_{\ell+1, \infty}$. Denote by $\boldsymbol{\gamma}_0 = \mathbf{Q}_2 \boldsymbol{\theta}_0$ the spline coefficients of s_{0, g_0} . Then we have the following decomposition: $\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0 = \tilde{\boldsymbol{\theta}}_\mu - \boldsymbol{\theta}_0 + \tilde{\boldsymbol{\theta}}_\epsilon$. Note that

$$\tilde{\boldsymbol{\theta}}_\mu - \boldsymbol{\theta}_0 = \mathbf{U}_{22} \mathbf{Q}_2^T \mathbf{B}^T \mathbf{H}_Z \mathbf{g}_0 - \boldsymbol{\theta}_0 = \mathbf{U}_{22} \mathbf{Q}_2^T \mathbf{B}^T \mathbf{H}_Z (\mathbf{g}_0 - \mathbf{B} \mathbf{Q}_2 \boldsymbol{\theta}_0) - \lambda \mathbf{U}_{22} \mathbf{Q}_2^T \mathbf{P} \mathbf{Q}_2 \boldsymbol{\theta}_0.$$

According to (2.10), for any \mathbf{a}

$$\mathbf{a}^T \mathbf{U}_{22}^{-1} \mathbf{a} = \mathbf{a}^T \mathbf{Q}_2^T (\mathbf{B}^T \mathbf{H}_Z \mathbf{B} + \lambda \mathbf{P}) \mathbf{Q}_2 \mathbf{a}.$$

Since \mathbf{H}_Z is idempotent, so its eigenvalues π_j is either 0 or 1. Without loss of generality we can arrange the eigenvalues in decreasing order so that $\pi_j = 1, j = 1, \dots, m$ and $\pi_j = 0, j = m+1, \dots, n$. Therefore, we have

$$\mathbf{a}^T (n\mathbf{U}_{22})^{-1} \mathbf{a} = \frac{1}{n} \sum_{j=1}^m \pi_j \mathbf{a}^T \mathbf{Q}_2^T \mathbf{B}^T \mathbf{e}_j \mathbf{e}_j^T \mathbf{B} \mathbf{Q}_2 \mathbf{a} + \frac{\lambda}{n} \mathbf{a}^T \mathbf{Q}_2^T \mathbf{P} \mathbf{Q}_2 \mathbf{a},$$

where \mathbf{e}_j be the indicator vector which is a zero vector except for an entry of one at position j . Using Markov's inequality, we have

$$\frac{\lambda}{n} \mathcal{E}_v \left(\sum_{\xi \in \mathcal{K}} a_\xi B_\xi \right) \leq \frac{\lambda}{n} \frac{C_1}{|\Delta|^2} C_2 \|\mathbf{a}\|^2.$$

Thus, by Conditions (C4) and (C5), $n\mathbf{a}^T\mathbf{U}_{22}\mathbf{a} \leq C|\Delta|^{-2}$. Next

$$\begin{aligned} & \|\mathbf{U}_{22}\mathbf{Q}_2^T\mathbf{B}^T\mathbf{H}_Z(\mathbf{g}_0 - \mathbf{BQ}_2\boldsymbol{\theta}_0)\| \leq C^{1/2}|\Delta|^{-1}n^{-1}\|\mathbf{B}^T\mathbf{H}_Z(\mathbf{g}_0 - \mathbf{BQ}_2\boldsymbol{\theta}_0)\| \\ & \leq C^{1/2}|\Delta|^{-1}n^{-1} \left[\sum_{\xi \in \mathcal{K}} \{\mathbf{B}_\xi^T\mathbf{H}_Z(\mathbf{g}_0 - \mathbf{BQ}_2\boldsymbol{\theta}_0)\}^2 \right]^{1/2} = O_P \left(\frac{F_2}{F_1} |\Delta|^\ell |g_0|_{\ell+1,\infty} \right), \end{aligned}$$

and

$$\lambda \|\mathbf{U}_{22}\mathbf{Q}_2^T\mathbf{PQ}_2\boldsymbol{\theta}_0\| \leq \frac{C\lambda}{n|\Delta|^4} \|s_{0,g_0}\|_{\mathcal{E}_v} \leq \frac{C\lambda}{n|\Delta|^4} \left(|g_0|_{2,\infty} + \frac{F_2}{F_1} |\Delta|^{\ell-1} |g_0|_{\ell+1,\infty} \right).$$

Thus,

$$\|\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\| = O_P \left\{ \frac{\lambda}{n|\Delta|^4} |g_0|_{2,\infty} + \left(1 + \frac{\lambda}{n|\Delta|^5} \right) \frac{F_2}{F_1} |\Delta|^\ell |g_0|_{\ell+1,\infty} \right\}.$$

For any $\boldsymbol{\alpha}$ with $\|\boldsymbol{\alpha}\| = 1$, we write $\boldsymbol{\alpha}^T \tilde{\boldsymbol{\theta}}_\epsilon = \sum_{i=1}^n \alpha_i \epsilon_i$ and $\alpha_i^2 = \boldsymbol{\alpha}^T \mathbf{U}_{22} \mathbf{Q}_2^T \mathbf{B}^T \mathbf{H}_Z \mathbf{B} \mathbf{Q}_2 \mathbf{U}_{22} \boldsymbol{\alpha}$. Following the same arguments as those in Lemma A.8, we have $\max_{1 \leq i \leq n} \alpha_i^2 = O_P(n^{-2} |\Delta|^{-4})$. Thus, $\|\tilde{\boldsymbol{\theta}}_\epsilon\| \leq |\Delta|^{-1} |\boldsymbol{\alpha}^T \tilde{\boldsymbol{\theta}}_\epsilon| = |\Delta|^{-1} |\sum_{i=1}^n \alpha_i \epsilon_i| = O_P(|\Delta|^{-2} n^{-1/2})$. Therefore,

$$\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\| = O_P \left\{ \frac{\lambda}{n|\Delta|^4} |g_0|_{2,\infty} + \left(1 + \frac{\lambda}{n|\Delta|^5} \right) \frac{F_2}{F_1} |\Delta|^\ell |g_0|_{\ell+1,\infty} + \frac{1}{\sqrt{n}|\Delta|^2} \right\}.$$

Observing that $\hat{g}(\mathbf{x}) = \mathbf{B}(\mathbf{x})\hat{\boldsymbol{\gamma}} = \mathbf{B}(\mathbf{x})\mathbf{Q}_2\hat{\boldsymbol{\theta}}$, we have

$$\|\hat{g} - g_0\|_{L^2} \leq \|\hat{g} - s_{0,g_0}\|_{L^2} + C \frac{F_2}{F_1} |\Delta|^{\ell+1} |g_0|_{\ell+1,\infty}.$$

According to Lemma A.1, we have.

$$\begin{aligned} \|\hat{g} - g_0\|_{L^2} & \leq C \left(|\Delta| \|\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0\| + \frac{F_2}{F_1} |\Delta|^{\ell+1} |g_0|_{\ell+1,\infty} \right) \\ & = O_P \left\{ \frac{\lambda}{n|\Delta|^3} |g_0|_{2,\infty} + \left(1 + \frac{\lambda}{n|\Delta|^5} \right) \frac{F_2}{F_1} |\Delta|^{\ell+1} |g_0|_{\ell+1,\infty} + \frac{1}{\sqrt{n}|\Delta|} \right\}. \end{aligned}$$

The proof is completed.

References

- Awanou, G. and Lai, M. J. and Wenston, P. (2005). The multivariate spline method for scattered data fitting and numerical solutions of partial differential equations. *Wavelets and splines: Athens 2005* 24–74.
- Chen, R. and Liang, H. and Wang, J. (2011). On determination of linear components in additive models. *Journal of Nonparametric Statistics* **23** 367–383.
- Craven, P. and Wahba, G. (1979). Smoothing noisy data with spline functions. *Numerische Mathematik* **31** 377–403.
- Eilers, P. (2006). *P-spline smoothing on difficult domains*. [online] Available at <http://www.statistik.lmu.de/sfb386/workshop/smcs2006/slides/eilers.pdf>
- Ettinger, B. and Guillas, S. and Lai, M. J. (2015). Bivariate splines for functional regression models with application to ozone forecasting. *Environmetrics* Accepted.
- Furrer, R. and Nychka, D. and Sainand, S. (2011). *Package ‘fields’. R package version 6.6.1..* [online] Available at <http://cran.r-project.org/web/packages/fields/fields.pdf>.
- Golitschek, M. and Schumaker, L. L. (2002). Bounds on projections onto bivariate polynomial spline spaces with stable local bases. *Constructive Approximation* **18** 241–254.
- Green, P. J. and Silverman, B. W. (1994). *Nonparametric regression and generalized linear models*. Chapman and Hall, London.
- Guillas, S. and Lai, M. J. (2010). Bivariate splines for spatial functional regression models. *Journal of Nonparametric Statistics* **22** 477–497.
- Härdle and W., Liang, H. and Gao, J. T. (2000). *Partially linear models*. Heidelberg: Springer Physica-Verlag.
- Huang, J. Z. and Zhang, L. and Zhou, L. (2007). Efficient estimation in marginal partially linear models for longitudinal/clustered data using splines. *Scandinavian Journal of Statistics* **34** 451–477.
- Lai, M. J. (2008). Multivariate splines for data fitting and approximation. *Conference Proceedings of the 12th Approximation Theory* 210–228.
- Lai, M. J. and Schumaker, L. L. (2007). *Spline functions on triangulations*. Cambridge University Press.

- Lai, M. J. and Schumaker, L. L. (1998). Approximation power of bivariate splines. *Advances in Computational Mathematics* **29** 595–623.
- Lai, M. J. and Wang, L. (2013). Bivariate penalized splines for regression. *Statistica Sinica* **23** 1399–1417.
- Li, Y. and Ruppert, D. (2008). On the asymptotics of penalized splines. *Biometrika* **95** 415–436.
- Liang, H. and Li, R. (2009). Variable selection for partially linear models with measurement errors. *Journal of the American Statistical Association* **104** 234–248.
- Liu, X. and Guillas, S. and Lai, M. J. (2015). Efficient spatial modeling using the SPDE approach with bivariate splines. *Journal of Computational and Graphical Statistics* Accepted.
- Liu, X. and Wang, L. and Liang, H. (2011). Estimation and variable selection for semiparametric additive partial linear models. *Statistica Sinica* **21** 1225–1248.
- Ma, S. and Song, Q. and Wang, L. (2013). Simultaneous variable selection and estimation in semiparametric modeling of longitudinal/clustered data. *Bernoulli* **19** 252–274.
- Ma, S. and Yang, L. (2011). Spline-backfitted kernel smoothing of partially linear additive model. *Journal of Statistical Planning and Inference* **141** 204–219.
- Marx, B. and Eilers, P. (2005) Multidimensional penalized signal regression. *Technometrics* **47** 13–22.
- Pace, R. K. and Barry, R. (1997) Sparse spatial autoregressions. *Statistics & Probability Letters* **33** 291–297.
- Persson, P. O. and Strang, G. (2002). A simple mesh generator in MATLAB. *SIAM Review* **64** 329–345.
- Ramsay, T. (2002). Spline smoothing over difficult regions. *Journal of the Royal Statistical Society, Series B* **64** 307–319.
- Sangalli, L. and Ramsay, J. and Ramsay, T. (2013). Spatial spline regression models. *Journal of the Royal Statistical Society, Series B* **75** 681–703.
- Shewchuk, J. R. (1996). *Applied Computational Geometry: Towards Geometric Engineering*. Springer-Verlag.

- Wahba, G. (1990). *Spline models for observational data*. SIAM Publications, Philadelphia. **136**, 2506-2534.
- Wang, L. and Liu, X. and Liang, H. and Carroll, R. (2011). Estimation and variable selection for generalized additive partial linear models. *Annals of Statistics* **39** 1827–1851
- Wang, H. and Ranalli, M. G. (2007). Low-rank smoothing splines on complicated domains. *Biometrics* **63** 209–217.
- Wang, L. and Xue, L. and Qu, A. and Liang, H. (2014). Estimation and model selection in generalized additive partial linear models for correlated data with diverging number of covariates. *Annals of Statistics* **42** 592–624.
- Wood, S. N. (2003). Thin plate regression splines. *Journal of the Royal Statistical Society, Series B* **65** 95–114.
- Wood, S. N. and Bravington, M. V. and Hedley, S. L. (2008). Soap film smoothing. *Journal of the Royal Statistical Society, Series B* **70** 931–955.
- Xiao, L. and Huang, L. and Schrack, J. A. and Ferrucci, L. and Zipunnikov, V. and Crainiceanu, C. (2015). Quantifying the life-time circadian rhythm of physical activity: a covariate-dependent functional approach. *Biostatistics* **16** 352–367.
- Xiao, L. and Li, Y. and Ruppert, D. (2013). Fast bivariate P-splines: the sandwich smoother. *Journal of the Royal Statistical Society, Series B* **75** 577–599.
- Zhang, H. and Cheng, G. and Liu, Y. (2011). Linear or nonlinear? Automatic structure discovery for partially linear models. *Journal of American Statistical Association* **106** 1099–1112.
- Zhou, L. and Pan, H. (2014). Smoothing noisy data for irregular regions using penalized bivariate splines on triangulations. *Computational Statistics* **29** 263–281.

Table 1: Root mean squared errors of the estimates in Example 1.

ρ	Method	β_1	β_2	σ_ε	$g(\cdot)$
0.0	KRIG	0.0850	0.0591	0.0497	0.4240
	TPS	0.0801	0.0541	0.0415	0.3268
	SOAP	0.0706	0.0511	0.0284	0.2164
	LFE	0.0712	0.0515	0.0305	0.1713
	QFE	0.0720	0.0520	0.0341	0.1800
	BPST (\triangle_1)	0.0705	0.0510	0.0285	0.1686
	BPST (\triangle_2)	0.0702	0.0509	0.0283	0.1669
0.7	KRIG	0.0867	0.0589	0.0454	0.4207
	TPS	0.0788	0.0565	0.0377	0.3224
	SOAP	0.0705	0.0545	0.0241	0.2016
	LFE	0.0746	0.0537	0.0288	0.1673
	QFE	0.0770	0.0533	0.0341	0.1765
	BPST (\triangle_1)	0.0706	0.0556	0.0237	0.1648
	BPST (\triangle_2)	0.0702	0.0549	0.0236	0.1624

Table 2: Standard error estimates of the BPST (\triangle_2) coefficients in Example 1.

ρ	Parameter	SE _{mc}	SE _{mean}	SE _{median}	SE _{mad}
0.0	β_1	0.0702	0.0660	0.0662	0.0032
	β_2	0.0501	0.0539	0.0536	0.0027
0.7	β_1	0.0723	0.0660	0.0661	0.0037
	β_2	0.0551	0.0540	0.0540	0.0028

Table 3: Root mean squared errors of the estimates in Example 2.

ρ	Method	β_1	β_2	σ_ε	MSPE
0.0	KRIG	0.0901	0.0754	0.3369	0.7700
	TPS	0.0835	0.0690	0.0656	0.7648
	SOAP	0.1106	0.0920	0.3223	0.9140
	LFE	0.1130	0.0930	0.4999	0.9649
	QFE	0.0940	0.0821	0.1223	0.8802
	BPST (\triangle_1)	0.0839	0.0699	0.0633	0.7647
	BPST (\triangle_2)	0.0799	0.0703	0.0425	0.7652
0.7	KRIG	0.0983	0.0733	0.3353	0.7552
	TPS	0.0906	0.0690	0.0612	0.8758
	SOAP	0.1269	0.0882	0.3404	0.9116
	LFE	0.1075	0.1067	0.4999	0.9496
	QFE	0.0988	0.0766	0.1316	0.8694
	BPST (\triangle_1)	0.0893	0.0692	0.0556	0.7574
	BPST (\triangle_2)	0.0891	0.0681	0.0450	0.7553

Table 4: Standard error estimates of the BPST (\triangle_2) coefficients in Example 2.

ρ	Parameter	SE _{mc}	SE _{mean}	SE _{median}	SE _{mad}
0.0	β_1	0.0803	0.0828	0.0821	0.0059
	β_2	0.0701	0.0673	0.0673	0.0054
0.7	β_1	0.0893	0.0893	0.0843	0.0046
	β_2	0.0679	0.0681	0.0683	0.0054

Table 5: 5-fold cross validation prediction errors for California housing data.

Method	OLS	KRIG	TPS	SOAP	LFE	QFE	BPST
Prediction Error (in log scale)	0.157	0.086	0.083	0.080	0.067	0.059	0.054

Table 6: Coefficients.

Variable	OLS	KRIG	TPS	SOAP	LFE	QFE	BPST
MedInc	0.202 (s.e. = 0.001)	0.085	0.094	0.148	0.139	0.122	0.114 (s.e. = 0.001)
(95% CI)	(0.200, 0.204)	NA	NA	NA	NA	NA	(0.112, 0.116)
log(Age)	0.165 (s.e. = 0.005)	-0.037	-0.039	-0.022	-0.038	-0.040	-0.035 (s.e. = 0.004)
(95% CI)	(0.155, 0.175)	NA	NA	NA	NA	NA	(-0.043, -0.027)
log(AveBedrms)	-0.037 (s.e. = 0.018)	0.068	0.083	0.133	0.112	0.100	0.095 (s.e. = 0.016)
(95% CI)	(-0.073, -0.001)	NA	NA	NA	NA	NA	(0.062, 0.128)
log(AveOccup)	-0.465 (s.e. = 0.022)	-0.063	-0.081	-0.276	-0.249	-0.169	-0.133 (s.e. = 0.018)
(95% CI)	(-0.509, -0.421)	NA	NA	NA	NA	NA	(-0.169, -0.097)
log(Hhd)	0.088 (s.e. = 0.004)	-0.004	-0.003	0.024	0.011	0.006	0.003 (s.e. = 0.003)
(95% CI)	(0.080, 0.096)	NA	NA	NA	NA	NA	(-0.002, 0.008)

[†] Estimates and confidence intervals (CI) are reported in the California house value study.

[‡] Note that the standard errors of $\hat{\beta}$ are not available for KRIG, TPS, SOAP, LFE and QFE methods, so the corresponding CIs are not available, and we leave them as “NA”.

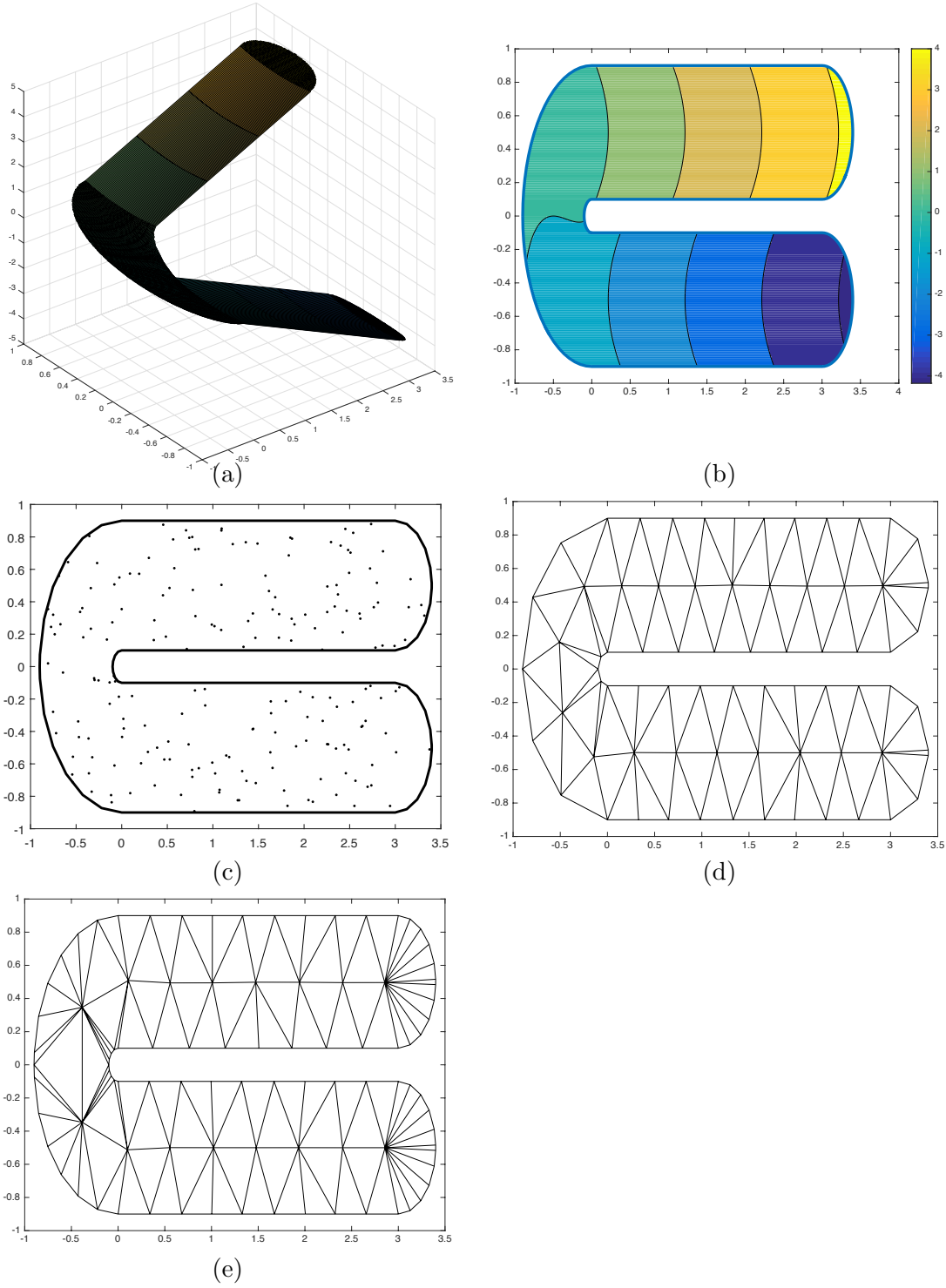


Figure 1: Example 1: (a) true function of $g(\cdot)$; (b) contour map of true function $g(\cdot)$; (c) sampled location points of replicate 1; (d) first triangulation (Δ_1) and (e) second triangulation (Δ_2) over the domain.

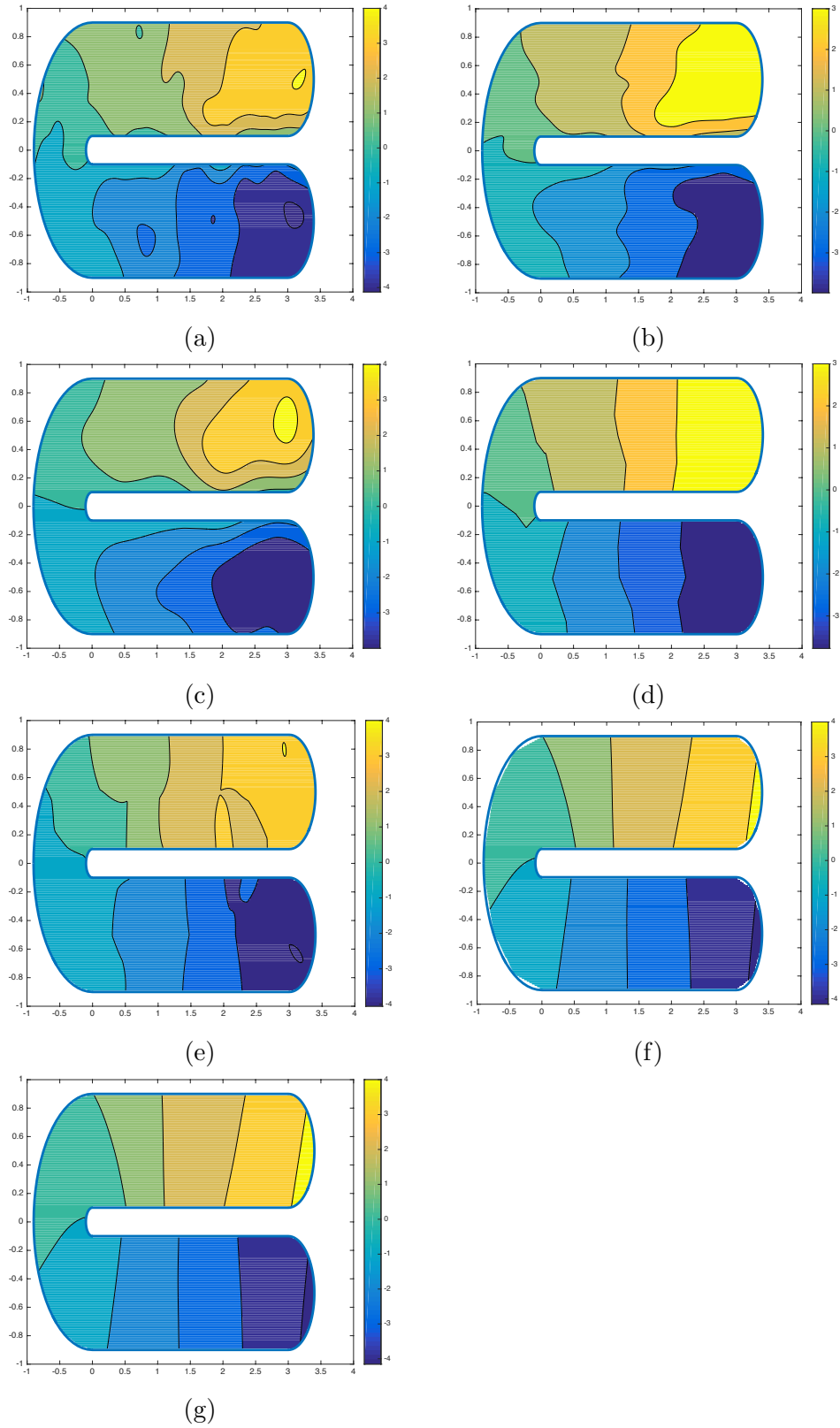


Figure 2: Contour maps for the estimators ($\rho = 0.0$) in Example 1: (a) KRIG; (b) TPS; (c) SOAP; (d) LFE; (e) QFE; (f) BPST (Δ_1) and (g) BPST (Δ_2).

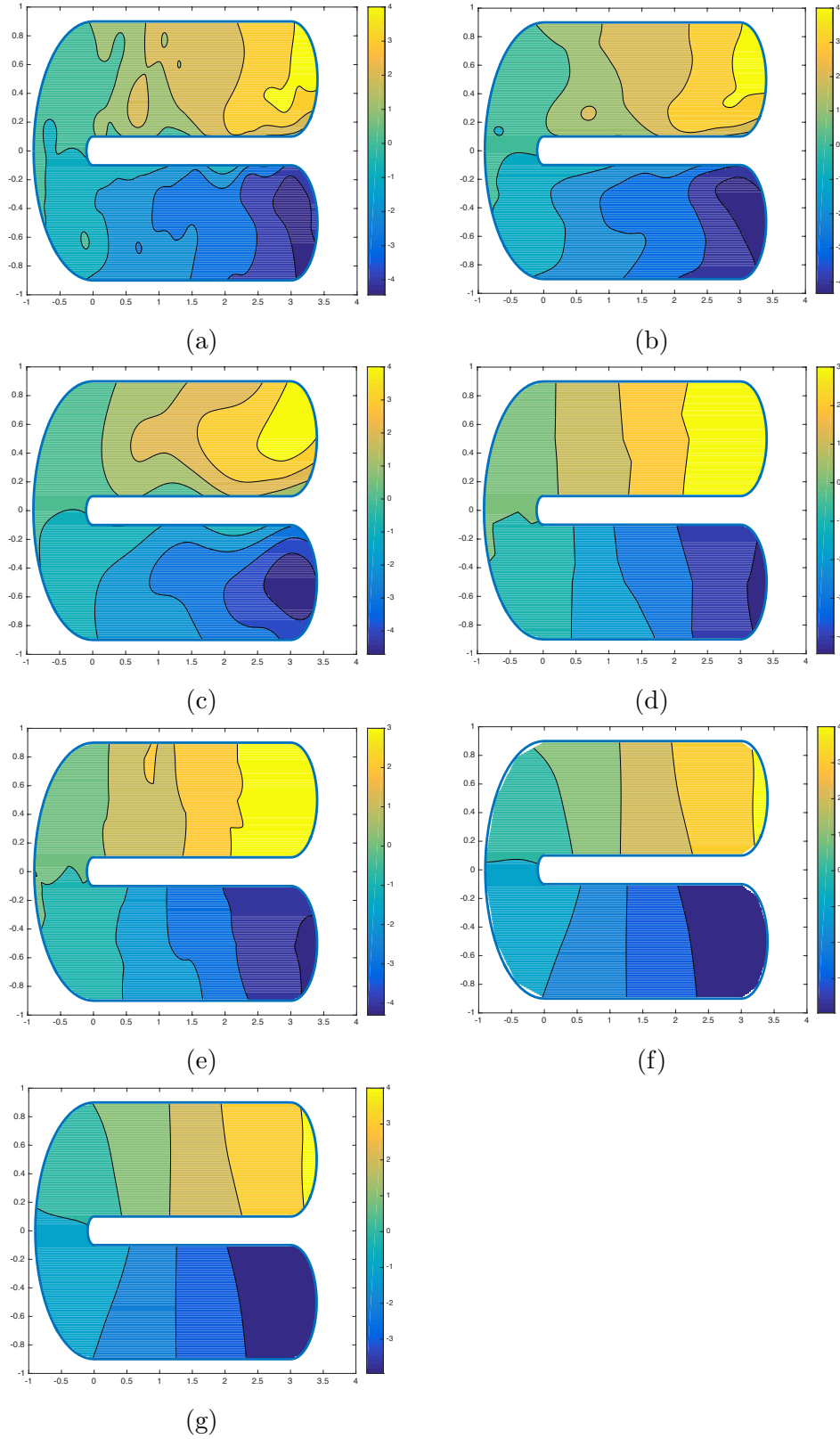


Figure 3: Contour maps for the estimators ($\rho = 0.7$) in Example 1: (a) KRIG; (b) TPS; (c) SOAP; (d) LFE; (e) QFE; (f) BPST (Δ_1) and (g) BPST (Δ_2).

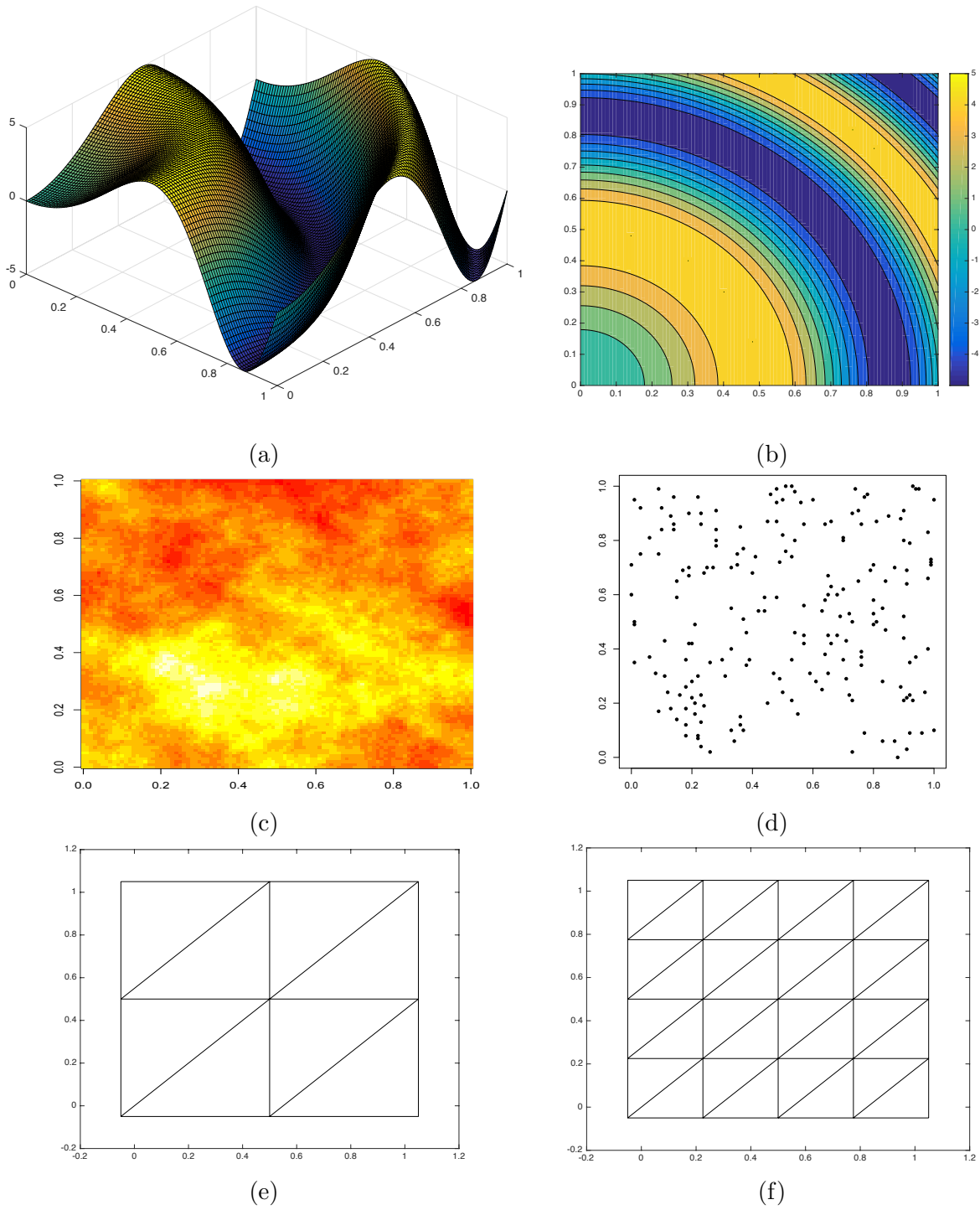


Figure 4: Example 2: (a) true function of $g(\cdot)$; (b) contour map of $g(\cdot)$; (c) gaussian random field $\xi(\cdot)$ and (d) sampled location points of replicate 1; (e) first triangulation (Δ_1) ; and (f) second triangulation (Δ_2) over the domain.

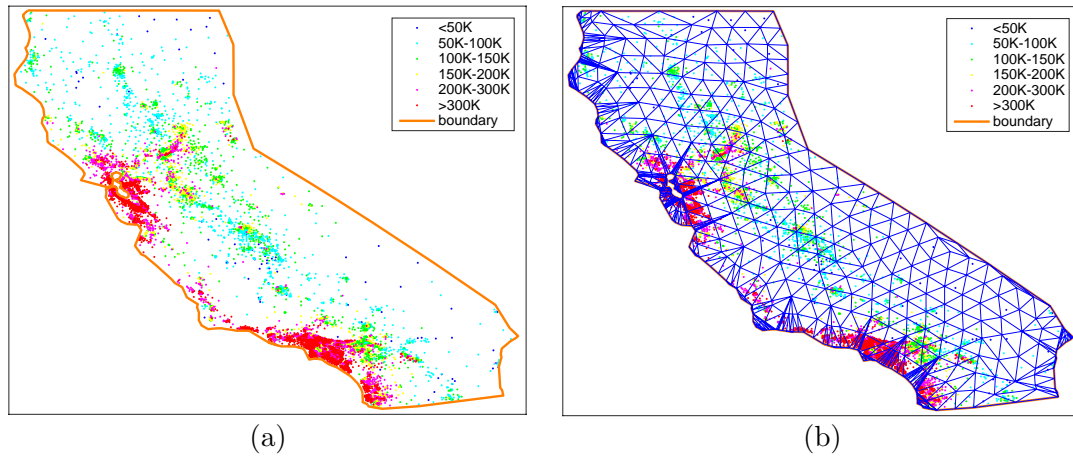


Figure 5: The median house values: (a) data location points (the colors of dots indicate different values of the houses); (b) domain triangulation.

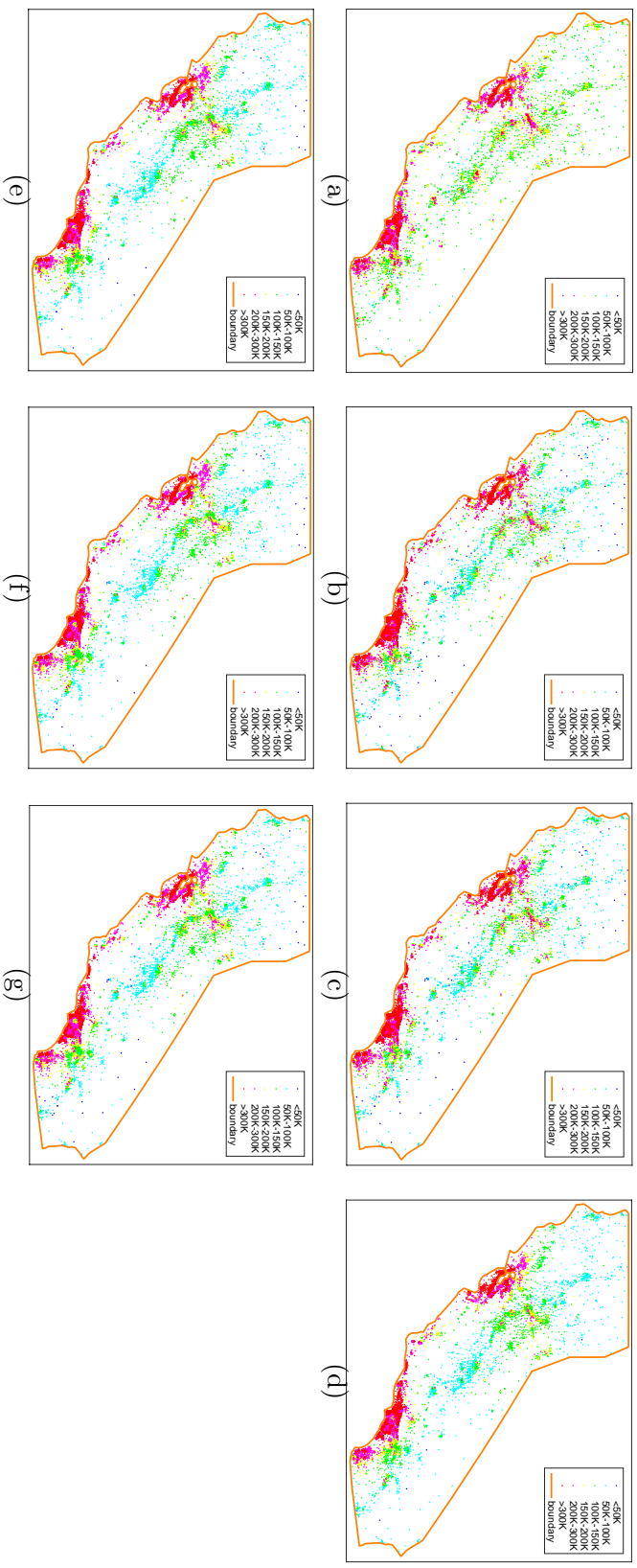


Figure 6: Scatter plots for the estimated house values using methods: (a) OLS; (b) KRIG; (c) TPS; (d) SOAP; (e) LFE; (f) QFE and (g) BPST.

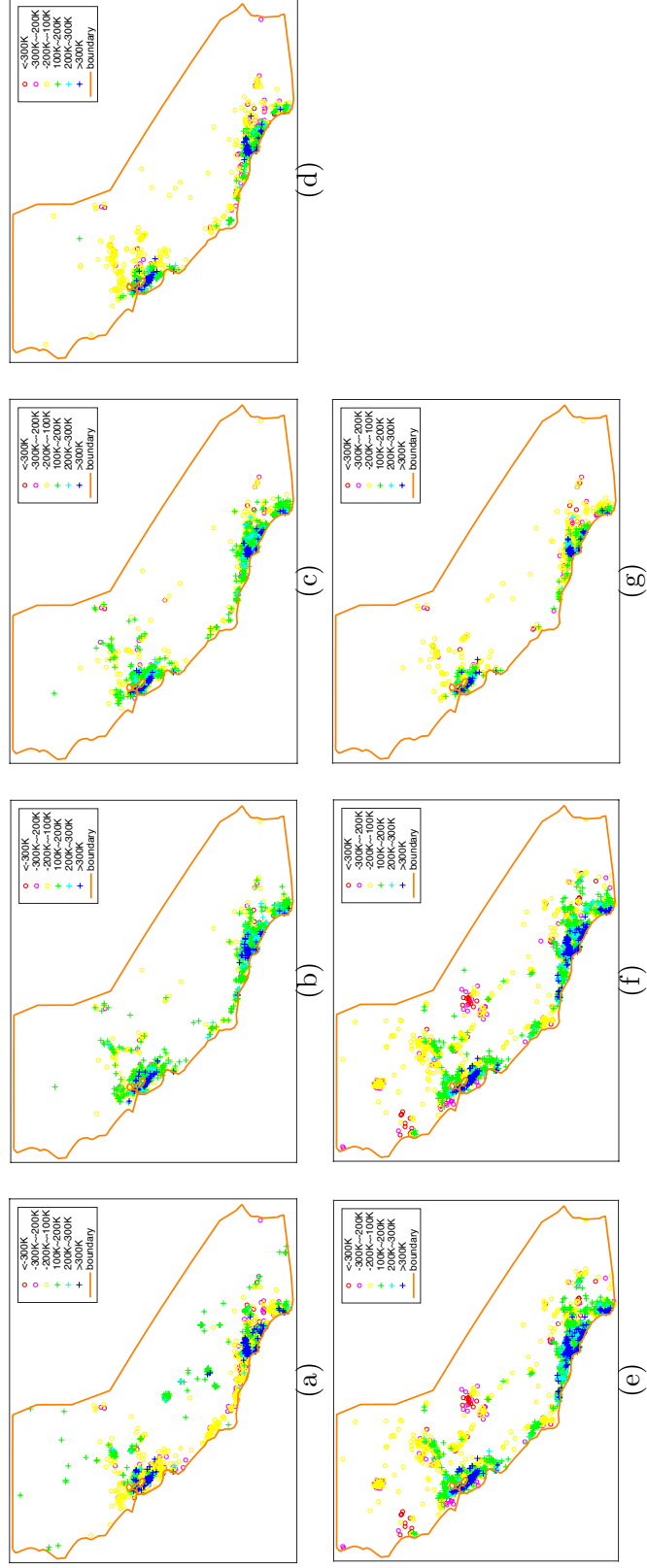


Figure 7: Scatter plots for the differences between the estimated house values and the observed house values using methods: (a) OLS; (b) KRIG; (c) TPS; (d) SOAP; (e) LFE; (f) QFE and (g) BPST.